1.0

4.5
5.0
5.6

2.8
3.2
3.6
4.0

2.5
2.2
2.0

1.1

1.8

1.25 1.4 1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

A MANUAL FOR ENCODING PROBABILITY DISTRIBUTIONS

Final Report

September 1978

By: Carl-Axel S. Staël von Holstein
James E. Matheson

Prepared for:

Defense Advanced Research Projects Agency
Arlington, Virginia

Under subcontract from
Decisions and Designs, Inc.
McLean, Virginia

SRI Project 7078

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. AD-AC72259	2.	3. Recipient's Accession No.

4. Title and Subtitle

Proability Encoding Manual

5. Report Date
September 1978

6.

7. Author(s) James E. Matheson
Carl-Axel S. Stael von Holstein

8. Performing Organization Rept. No.
7078

9. Performing Organization Name and Address

SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

10. Project/Task/Work Unit No.

11. Contract(C) or Grant(G) No.

(C) ONR-N00014-78-C-0100

(G)

12. Sponsoring Organization Name and Address

Advanced Research Projects Agency
Cybernetics Technology Division
1400 Wilson Blvd.
Arlington, VA. 22209

13. Type of Report & Period Covered

Technical

14.

15. Supplementary Notes

16. Abstract (Limit: 200 words)

Accounting for the uncertainty in most decision problems requires the application of the knowledge of one or more experts. In decision analysis, a process of probability encoding is used to convert specialized or general knowledge into probability distributions to represent the judgement of these experts. This manual presents the probability encoding methods that have proved to be most effective in providing a precise language for communication about uncertainty.

The interview process described herein has been found to be an efficient method for eliciting the uncertainty in a person's information. The manual details the careful interaction between the interviewer (decision analyst) and the subject whose judgment is encoded. The encoding process not only provides the user with an awareness of the many problems that may arise in an encoding situation, but also improves the subject's awareness of his state of information.

17. Document Analysis   a. Descriptors

b. Identifiers/Open-Ended Terms

Decision Analysis, Probability Encoding, Probability Assessment

c. COSATI Field/Group

18. Availability Statement:	19. Security Class (This Report)	21. No. of Pages
Release Unlimited	Unclassified	236
	20. Security Class (This Page) Unclassified	22. Price

(See ANSI-Z39.18) See Instructions on Reverse

OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce

DO NOT PRINT THESE INSTRUCTIONS AS A PAGE IN A REPORT

INSTRUCTIONS

Optional Form 272, Report Documentation Page is based on Guidelines for Format and Production of Scientific and Technical Reports, ANSI Z39.18–1974 available from American National Standards Institute, 1430 Broadway, New York, New York 10018. Each separately bound report—for example, each volume in a multivolume set—shall have its unique Report Documentation Page.

1. **Report Number.** Each individually bound report shall carry a unique alphanumeric designation assigned by the performing organization or provided by the sponsoring organization in accordance with American National Standard ANSI Z39.23–1974, Technical Report Number (STRN). For registration of report code, contact NTIS Report Number Clearinghouse, Springfield, VA 22161. Use uppercase letters, Arabic numerals, slashes, and hyphens only, as in the following examples: FASEB/NS–75/87 and FAA/RD–75/09.

2. Leave blank.

3. **Recipient's Accession Number.** Reserved for use by each report recipient.

4. **Title and Subtitle.** Title should indicate clearly and briefly the subject coverage of the report, subordinate subtitle to the main title. When a report is prepared in more than one volume, repeat the primary title, add volume number and include subtitle for the specific volume.

5. **Report Date.** Each report shall carry a date indicating at least month and year. Indicate the basis on which it was selected (e.g., date of issue, date of approval, date of preparation, date published).

6. **Sponsoring Agency Code.** Leave blank.

7. **Author(s).** Give name(s) in conventional order (e.g., John R. Doe, or J. Robert Doe). List author's affiliation if it differs from the performing organization.

8. **Performing Organization Report Number.** Insert if performing organization wishes to assign this number.

9. **Performing Organization Name and Mailing Address.** Give name, street, city, state, and ZIP code. List no more than two levels of an organizational hierarchy. Display the name of the organization exactly as it should appear in Government indexes such as Government Reports Announcements & Index (GRA & I).

10. **Project/Task/Work Unit Number.** Use the project, task and work unit numbers under which the report was prepared.

11. **Contract/Grant Number.** Insert contract or grant number under which report was prepared.

12. **Sponsoring Agency Name and Mailing Address.** Include ZIP code. Cite main sponsors.

13. **Type of Report and Period Covered.** State interim, final, etc., and, if applicable, inclusive dates.

14. **Performing Organization Code.** Leave blank.

15. **Supplementary Notes.** Enter information not included elsewhere but useful, such as: Prepared in cooperation with . . . Translation of . . . Presented at conference of . . . To be published in . . . When a report is revised, include a statement whether the new report supersedes or supplements the older report.

16. **Abstract.** Include a brief (200 words or less) factual summary of the most significant information contained in the report. If the report contains a significant bibliography or literature survey, mention it here.

17. **Document Analysis.** (a). *Descriptors.* Select from the *Thesaurus of Engineering and Scientific Terms* the proper authorized terms that identify the major concept of the research and are sufficiently specific and precise to be used as index entries for cataloging.

    (b). *Identifiers and Open-Ended Terms.* Use identifiers for project names, code names, equipment designators, etc. Use open-ended terms written in descriptor form for those subjects for which no descriptor exists.

    (c). *COSATI Field/Group.* Field and Group assignments are to be taken from the 1964 COSATI Subject Category List. Since the majority of documents are multidisciplinary in nature, the primary Field/Group assignment(s) will be the specific discipline, area of human endeavor, or type of physical object. The application(s) will be cross-referenced with secondary Field/Group assignments that will follow the primary posting(s).

18. **Distribution Statement.** Denote public releasability, for example "Release unlimited", or limitation for reasons other than security. Cite any availability to the public, with address, order number and price, if known.

19. & 20. **Security Classification.** Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED).

21. **Number of pages.** Insert the total number of pages, including introductory pages, but excluding distribution list, if any.

22. **Price.** Enter price in paper copy (PC) and/or microfiche (MF) if known.

SUMMARY

The uncertainty inherent in most decision problems usually requires the application of the knowledge of one or more experts. Decision analysis uses a process of probability encoding to convert this specialized or general knowledge into probability distributions that represent judgment of these experts. Probability encoding can be applied to uncertainties in any decision-making situation.

The purpose of this manual is to present the methods of probability encoding that the Decision Analysis Group (DAG) at SRI International currently uses. The manual focuses on the probability encoding methods that have been developed through years of experience in a wide range of decision analysis applications and through evidence obtained from experiments performed by DAG and others (notably Professors Daniel Kahneman and Amos Tversky). Hence, this manual presents those methods that have proved to be most effective in providing a precise language for communication about uncertainty.

Experience has shown that the interview is the most effective method of encoding a probability distribution. Through the interview process (described in Chapter 5), the uncertainty in a person's information is elicited. The resulting probability distribution is derived from careful interaction between the interviewer (decision analyst) and the subject (the person whose judgment is encoded). Because subjects consciously or unconsciously have biases in their responses, these responses do not always accurately reflect the underlying judgments. It is therefore important that the interviewer be able to detect and help to reduce or eliminate such biases through the way the interview is structured and the questions are asked. (Chapter 3 discusses biases in detail.)

This manual is based on an interpretation fundamental to the total decision analysis philosophy that probability represents "an encoding of information." Different state of knowledge can be translated into probabilities that can be compared. For example, because various people are likely to have different states of information, two persons can assign a different probability to the same uncertain quantity. Moreover, if a person receives new relevant information, he is likely to revise his probability assignment. Thus, in this frame of reference, information can be defined as anything that causes a person to change his probability assignment.

Probability encoding is a major element of decision analysis. Although probability encoding usually is performed in the ontext of a decision problem, its general benefit extends beyond the analysis of specific decisions. It improves a subject's awareness of his state of information and provides a clear method for communication and inference about uncertainty.

A decision analysis usually includes three phases--the deterministic, probabilistic, and informational phases. In the deterministic phase, the decision problem is structured by defining relevant variables, characterizing their relationships in formal models, and assigning values to possible outcomes. The importance of the different variables is measured through sensitivity analysis.

During the probabilistic phase, uncertainty is explicitly incorporated into the analysis by assigning probability distributions to the important variables. These distributions are obtained by encoding the judgment of individuals knowledgeable about the problem. These judgments are processed using the models developed in the deterministic phase and are transformed into a probability distribution that expresses the uncertainty about the final outcome. After the decision maker's

attitude toward risk has been evaluated and taken into account, the best alternative in the face of uncertainty is established.

In the informational phase, the economic value of information is determined by calculating the worth of reducing uncertainty about each of the important variables in the problem. The value of additional information can then be compared with the cost of obtaining it. If the gathering of additional information is profitable, the three phases are repeated again. The analysis is completed when further analysis or information gathering is no longer profitable.

Throughout the course of any decision analysis, the procedure focuses on the decision and the decision maker. Expanding an analysis is considered valuable only if it helps the decision maker choose among available alternatives.

The decision maker is the person (or group of persons) who has responsibility for the decisions under consideration. It follows that a decision analysis must be based on the decision maker's beliefs and preferences. He may be willing to designate some other person or persons as his expert(s) for encoding the uncertainty in a particular variable if he feels that the expert has a more relevant information base. The decision maker can then either accept the expert's information as his input to the analysis or modify it to incorporate his own judgment.

The time and effort expended on probability encoding depend on the importance of the decision problem, the importance of the variable under consideration, and so on. Sometimes it is useful to spend a great deal of time and effort on encoding probabilities. For such important variables, the judgments of more than one person are often encoded. In other situations, only a brief encoding effort may be necessary.

The manual focuses on the probability encoding methods that have been found to be effective rather than on a presentation of a comprehensive overview of the entire decision analysis process. Therefore, the manual is intended for the use of interviewers who are knowledgeable in decision analysis. Although the reader should not expect to become a complete expert in probability encoding, the manual should provide the user with an awareness of the many problems that may arise in an encoding situation.

PREFACE

Uncertainty plays a major role in most decision problems. Probability distributions are used in decision analysis to describe formally the uncertainty regarding important inputs to the analysis. These distributions often represent the judgment of experts. The process of developing a formal distribution from a subject's judgment is called probability encoding. This manual focuses on encoding processes that are conducted by trained interviewers. The user is assumed to have some familiarity with decision analysis and the role of probability encoding in decision analysis.

The manual is based on many years of practical experience with probabili·y encoding in decision analysis applications within the Decision Analysis Group at SRI International (formerly Stanford Research Institute). We have had many discussions of the subject with our colleagues, in particular with Dr. Carl S. Spetzler, who was a coauthor of an earlier version of this manual, and with Mr. Ramon Zamora, and we are grateful for their valuable comments. Drs. Daniel Kahneman and Amos Tversky have served as consultants to SRI on this subject. Most of the material in Chapter 3 is based on their work. They also prepared the report presented here as Appendix A for SRI in 1975, and it summarizes much of their research in this area.

CONTENTS

DISTRIBUTION LIST

ILLUSTRATIONS

TABLE

ACKNOWLEDGMENT

GLOSSARY

The glossary of commonly used terms in this manual is provided for readers and users who may have limited experience with either decision analysis or probability encoding. The list is also provided to help clarify terms that are perhaps used in slightly different ways by the authors.

Analyst
The individual who has the responsibility for a decision analysis. In probability encoding, he may be the interviewer.

Assessment (of Probability)
See **Probability Encoding.**

Assignment (of Probability)
Expressing the judgment regarding the uncertainty of an event in terms of a probability. See also **Probability Encoding.**

Bias
A conscious or subconscious discrepancy between a subject's response and an accurate description of his underlying knowledge. See Section 3.2.

Clairvoyance Test
To ask whether a clairvoyant could reveal the value of an uncertain quantity by specifying a single number without asking questions for clarification.

Cumulative (Probability) Distribution
A function (curve) that for each possible value of an uncertain quantity gives the probability that the quantity will attain at most that value.

Decision
An essentially irrevocable allocation of resources.

Decision Analysis
The discipline concerned with the formal analysis of decision problems.

Decision Maker
The individual (or organizational entity) who has the responsibility and authority to commit the resources that constitute a decision.

Decision Variable	A system variable under the direct control of the decision maker.
Density Function	Also Probability Density Function. A function (curve) describing the relative likelihood of the occurrence of the possible values of an uncertain quantity.
Discrete Distribution	A probability distribution for an uncertain quantity with a limited (or countable) number of outcomes.
Elicitation	See **Probability Encoding.**
Encoding	See **Probability Encoding.**
Expert	A person knowledgeable about an uncertain quantity.
Fractile	The value of an uncertain quantity that corresponds to a given probability level.
Histogram	A density function in bar graph form. The height of any bar represents the probability of obtaining a value within the corresponding interval.
Interviewer	The person conducting a probability encoding.
Joint Distribution	A probability distribution for two or more uncertain quantities.
Mass Function	A density function for a discrete distribution. It shows the probability for each possible outcome.
Median	The value of an uncertain quantity for which the cumulative probability is 0.5.
Modes of Judgment	Conscious and unconscious procedures (heuristics) that people use in making intuitive probabilistic estimates.
Probability	A number between 0 and 1 (inclusively) representing the degree of belief a person attaches to the occurrence of an event.

Probability Distribution	A representation of the uncertainty in someone's judgment about an uncertain quantity. See also **Cumulative Distribution, Density Function, Histogram, Mass Function.**
Probability Encoding	The process by which a person's judgment regarding an uncertain quantity is characterized by a probability distribution. Some terms in the literature that are used synonymously to Encoding: Assessment, Assignment (used for single events in this manual), and Elicitation.
Quartile	The value of an uncertain quantity for which the cumulative probability is 0.25 or 0.75.
Reference Process	A generator of uncertain quantities with which the subject has good familiarity.
State Variable	A system variable that is controlled by the environment.
Subject	The person whose judgment is encoded (generally expected to be an expert).
System Variable	Those variables (state and decision) on which the outcomes depend.
Uncertain Quantity	A quantity whose value is uncertain. Generally termed state variables in decision analysis applications.

1 INTRODUCTION

Decision analysis often requires that the knowledge of several different experts be combined. The universal code which permits different kinds of knowledge to be combined in a consistent manner is probability. The process whereby the specialized or general knowledge of an expert is converted to a probability assignment is called probability encoding. Questions related to the process of probability encoding have been asked and studied since the early 1950s, primarily by psychologists but also by economists, statisticians, and decision theorists. The greater part of the published research has been restricted to laboratory experiments with simple paradigms, and the results cannot be easily extrapolated to real-world applications of probability encoding for decision analysis. The main reason for the inapplicability is that the primary issues and problems that are found when a subject is interviewed to incorporate his knowledge in a decision analysis have not been tackled in the laboratory. There are some exceptions, such as the work by Kahneman and Tversky reported in Appendix A. It is our firm belief that probability encoding procedures need to be based on practical experience and experimentation. This manual represents our current understanding of the field based on our experience and that of our colleagues in numerous encoding sessions conducted as parts of practical decision analyses as well as on experimental evidence in the literature.

Probability encoding can be applied to any field of uncertainty. In fact, applications range from military strategy to exploration of Mars to business decision-making. Examples throughout the manual are drawn from a variety of fields.

1.1 Probability

Probability encoding is the process of extracting and quantifying individual judgment about a well-defined quantity (which though well-defined is uncertain). The purpose of probability encoding is to provide a precise language for communication about uncertainty. This aspect of probability should be distinguished from probability calculus, which is concerned with technical manipulations of probabilities. Although the mathematical results of probability calculus are seldom disputed, the interpretation of probability has been the subject of extensive debate.[*]

This manual uses an interpretation that is fundamental to the whole decision analysis philosophy. In this view, probability represents "an encoding of information." Probability is a language into which different states of knowledge may be translated and thereby compared or used in a common way. In that various people are likely to have different information, two persons can make different probability assignments to the same uncertain quantity. Furthermore, a person is likely to revise a probability assignment if he receives new and relevant information. Indeed, the concept "information" in this frame of reference may be defined as anything that causes a person to change a probability assignment; whether a message contains information depends upon what you already know. If it does not alter your probability assignment, you may conclude that the message contains no information for you, i.e., you knew it already.

A probability distribution must meet these three conditions:

(1) The probability of any event is a number between zero and one (inclusively).

(2) The probability of an event that is certain to occur is one.

---

[*]For an exposition of various interpretations see B. de Finetti, "Probability: Interpretations," in *International Encyclopedia of the Social Sciences*, D. L. Sills (ed.), vol. 12, pp. 496-505 (Macmillan and Free Press, New York, 1968).

(3) If there are two events that cannot occur simultaneously,
then the probability that one event or the other will
occur is equal to the sum of the probabilities for each
of the two events.

No other constraints restrict the probability assignment.

A probability distribution can be represented in two different
ways. The cumulative probability distribution (cumulative distribution,
for short) is one way; an example is shown in Figure 1(a). The possible
outcomes for the uncertain quantity are expressed on a value scale (the
horizontal axis); the uncertainty in the outcomes is expressed on a
probability scale (the vertical axis). Each number on the value scale
has a corresponding probability level, which is interpreted as the
probability that the resolved value of the uncertain quantity will be
less than or equal to that number. For example, in Figure 1(a), the
probability that the demand will not exceed 35,000 machines is assigned
as 0.65.

An alternative way of describing the distribution is through the
density function. The density function has the property that the area
under the curve between any two points is equal to the probability that
the resolved value will fall between those points. (The total area
under the curve is therefore one.) The height of the density function
at a particular point thus indicates the relative probability that the
resolved value will fall in the vicinity of that point. The density
function corresponding to the cumulative distribution of Figure 1(a) is
shown in Figure 1(b).

The cumulative distribution and the density function are two
equivalent ways of describing a probability distribution. The area
under the density function between any two points is always equal to the
difference in values of the cumulative distribution for the same points.
The steeper the cumulative distribution is, the higher the density
function will be. The relationship between the cumulative distribution
and the density function will be further illustrated in Section 5.5.

3
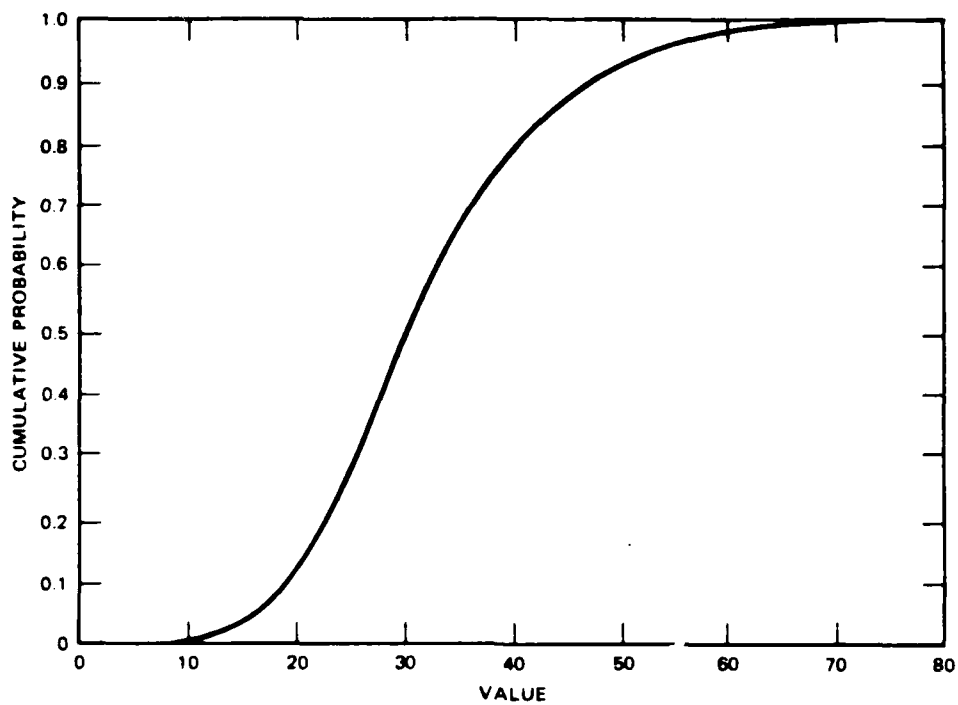
(a) CUMULATIVE DISTRIBUTION
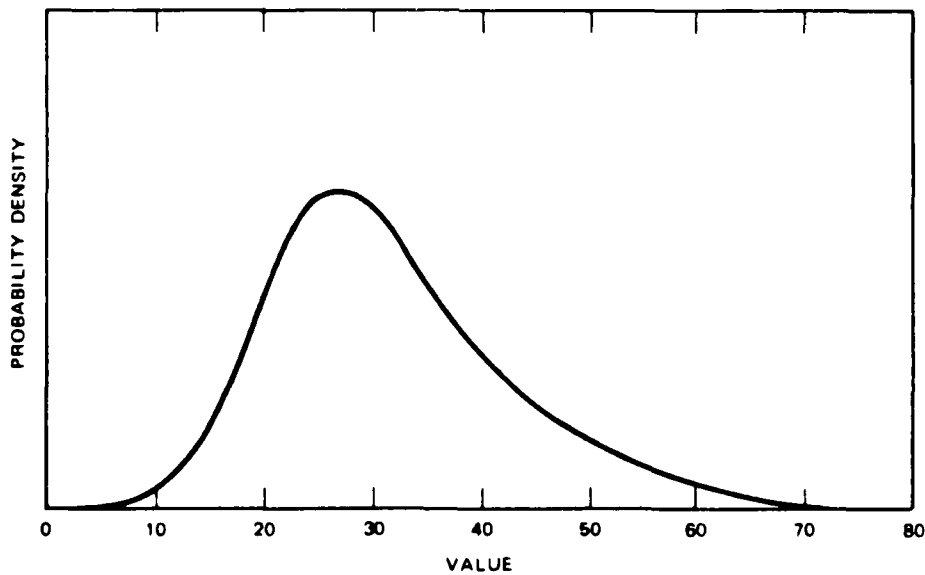
(b) DENSITY FUNCTION

FIGURE 1   TWO REPRESENTATIONS OF A PROBABILITY DISTRIBUTION

4

1.2 The Decision Analysis Framework

Probability encoding is a major element of decision analysis. It is usually performed in the context of *a specific decision problem.* However, the general benefit of probability encoding extends beyond the analysis of specific decisions. It improves a subject's awareness of his state of information and provides a clear means for communication and inference about uncertainty. A brief overview of decision analysis *is given below to provide a frame of reference.* *

A decision analysis usually includes three phases--the deterministic, probabilistic, and informational phases. In the <u>deterministic phase</u>, the decision problem is structured by defining *relevant* variables, characterizing their relationships in formal models, and assigning values to possible outcomes. The importance of the different variables is measured through sensitivity analysis.

During the <u>probabilistic phase</u>, uncertainty is explicitly incorporated into the analysis by assigning probability distributions to the important variables.** These distributions are obtained by encoding the judgment of individuals knowledgeable about the problem. These judgments are processed using the models developed in the deterministic phase and are transformed into a probability distribution that expresses the uncertainty about the final outcome. After the decision maker's attitude toward risk has been evaluated and taken into account, the best alternative in the face of uncertainty is established.

---

*A more extensive discussion of decision analysis will be found in R. A. Howard, "Decision Analysis: Applied Decision Theory," in <u>Proceedings of the Fourth International Conference on Operational Research</u>, D. B. Hertz and J. Melese, eds., pp. 55-71 (John Wiley and Sons, New York, 1968), and in R. A. Howard, "The Foundations of Decision Analysis," <u>IEEE Transactions on Systems Science and Cybernetics</u>, Vol. SSC-4, pp. 211-219 (1968).

**We use 'uncertain quantity' rather than 'variable' when discussing probability encoding in this manual. However, we will use variable in the context of decision analysis methodology to conform with general decision analysis terminology.

In the _informational phase_, the economic value of information is determined by calculating the worth of reducing uncertainty about each of the important variables in the problem. The value of additional information can then be compared with the cost of obtaining it. If the gathering of additional information is profitable, the three phases are repeated again. The analysis is completed when further analysis or information gathering is no longer profitable.

Throughout the course of any decision analysis, the procedure focuses on the decision and the decision maker. Expanding an analysis is considered valuable only if it helps the decision maker choose among available alternatives.

The decision maker is the person (or group of persons) who has responsibility for the decisions under consideration. It follows that a decision analysis must be based on the decision maker's beliefs and preferences. He may be willing to designate some other person or persons as his expert(s) for encoding the uncertainty in a particular variable if he feels that the expert has a more relevant information base. The decision maker can then either accept the expert's information as his input to the analysis or modify it to incorporate his own judgment.

The time and effort expended on probability encoding depends on the importance of the decision problem, the importance of the variable under consideration, and so on. Sometimes it is useful to spend a great deal of time and effort on encoding probabilities. For such important variables, the judgments of more than one person are often encoded. In other situations, only a brief encoding effort may be necessary.

1.3 Purpose of the Manual

Probability encoding is the quantification in terms of a probability distribution of someone's judgment (state of information) regarding an uncertain quantity. This is, first of all, a matter of bringing a person's information to his conscious attention and making

6

him realize what he does and does not know. We have found an interview process to be the most effective way of encoding a probability distribution. The distribution is the result of careful interaction between the interviewer (the analyst) and the subject (the person whose judgment is encoded).

Experience has shown that subjects consciously or unconsciously often make responses that are biased in some way--i.e., the responses do not accurately reflect the subjects' underlying judgments. It is therefore important for the interviewer to be able to detect and correct such biases and to reduce or eliminate bias through the way the interview is structured and questions are asked.

This manual describes probability encoding methods currently used by members of the Decision Analysis Group at SRI International. These methods are based on years of experience with probability encoding in applications of decision analysis as well as evidence from experiments performed by the Decision Analysis Group and others (notably Professors Daniel Kahneman and Amos Tversky). The manual is not intended to be a *comprehensive overview of the whole problem*, but rather to be a presentation of methods that have been found effective. The manual is written for interviewers who are familiar with decision analysis. An interviewer should not expect to become a complete expert in probability encoding from reading the manual, but he should become more aware of many problems that may arise in an actual encoding situation.

1.4 Outline of the Manual

The next three chapters provide the background knowledge necessary for the interview process. Chapter 2 deals with the interaction between modeling and encoding. Modeling and encoding represent two forms of quantification of judgment, and a balance has to be found in each problem. Chapter 3 discusses biases that may occur when people are making judgments of uncertainties. Chapter 4 describes a variety of probability encoding techniques. Chapter 5, which can be considered the central part of the manual, describes the interview process. The

7

process is usually based on a variety of encoding techniques, and the interviewer should be alert for any biases that might be introduced or corrected by the use of such techniques. Chapter 6 provides an example of an encoding session that illustrates many of the concepts discussed in the preceding chapter. Chapter 7 presents miscellaneous topics that have been collected and provided as additional background material. An annotated bibliography on probability encoding is provided.

The manual has 4 appendices. Appendix A supplements Chapters 3 and 5 with a more detailed discussion of biases and corrective procedures. Appendix B provides a step-by-step presentation of the interview process, which is illustrated by sample questions and answers. Appendix C is an encoding interview form that follows the interview process described in Appendix B. Appendix D contains a sample session with an interactive computer interview.

2 MODELING AND ENCODING

2.1 Decision and State Variables

A decision analysis model includes two kinds of input variables: decision variables and state variables. The two must be carefully distinguished from one another because the decision maker can choose the values of the decision variables, whereas the values of the state variables are beyond his control. Thus, it is only meaningful to discuss encoding with respect to state variables.

Some variables may at first seem difficult to classify as decision or state variables. This difficulty, however, may be resolved by further structuring of the problem. For example, price may be separated into a controllable price strategy and the uncertain market response. A similar problem can arise when variables interact. For example, in new system decisions, the variables of development time, program cost, and system performance are closely related. To solve this problem, one or two variables can be selected as decision variables, and the others can then be treated as state variables. Before any probability encoding can begin, every decision problem must be carefully structured so that it is clear which variables are best considered decision variables and which are state variables. Often this separation is most easily achieved by redefinition of the variables, as will be illustrated later in this section.

2.2 The Balance Between Modeling and Encoding

During the modeling stage of a decision problem, there is always the question of whether to directly encode the uncertainty in an important variable or to further model the problem. At one extreme, it is conceivable that the final worth or profit of an alternative could be encoded directly, thus bypassing a need for examination of underlying

9

variables. Generally, however, a probability distribution for final worth is more easily reached and engenders more confidence if a model is constructed that relates final worth to underlying variables. Modeling efforts tend to be most effective and most economical if they start with a gross model that is successively refined. A model should be refined only as long as the cost of each addition provides at least comparable improvement in information. The criterion for how much information is needed depends on how significantly the information bears on the decision at hand.

A decision whether or not to launch a new product development illustrates how the degree of modeling must be adapted to suit the problem at hand. Naturally, one of the most important factors in decisions regarding new introductions is the size of the market for the product. The simplest model might consider the market as a whole and define it by total market potential, company market share, and average growth rate. This is obviously a crude description, but in many cases it may be sufficient for the decision at hand. For a product with many potential markets with different characteristics, it may be necessary to expand the model to describe some of the markets separately. However, sensitivity analysis often shows that even though the markets are defined explicitly in the model, not all of them need be considered for probability encoding. In a recent decision analysis, probability distributions were encoded for the total international market and for three major domestic applications. At the same time, the remaining domestic applications were combined into one variable that, although uncertain, did not need to be considered probabilistically. In some situations, though, the complexity and importance of the problem necessitates a complicated model structure. For example, a recent study of a plant decision in the oil industry required a thorough model of the whole industry in order to assess the company's own market and price situation.

A choice between additional modeling and encoding may also need to be reconsidered during the encoding process. The subject may find it

easier to think about the problem in terms of a different structure. He may also reveal biases during the interview that can be counteracted by further structuring of the problem.

2.3 Guidelines for Preparing To Encode Uncertain Quantities

We offer the following list of principles as an aid in defining and structuring any variables whose uncertainty is to be encoded. From our experience, violating these principles leads to problems in the probability encoding process.

- Choose only uncertain quantities that are important to the decision, as determined by a sensitivity analysis. Be prepared to explain to the subject why the quantity is important to the decision. This demonstrates the relevance of the encoding process and is essential in gaining the subject's full cooperation.

- Define the quantity as an unambiguous state variable. If the subject believes that his decision can affect the outcome of the quantity to some extent, then the problem needs restructuring to eliminate this effect.

- Structure the quantity carefully. The subject may think of the quantity as conditional on other quantities. If so, conditionalities should be considered consciously and be incorporated into the model structure because it is difficult for human minds to deal effectively with combinations of uncertain quantities. For example, a major consideration for someone forecasting the sales of a new product might be whether or not the main competitor will develop a similar product. The encoding might then be facilitated by making two separate probability assignments—one for the case where the competitor exists and one where he does not. Mental acrobatics should be minimized.

- Define the quantity clearly. A good test is to ask whether a clairvoyant could reveal the value of the quantity by specifying a single number without requesting clarification. For example, it is not meaningful to ask for the "price of wheat in 1980," because the clairvoyant would need to know the quantity, kind of wheat, at what date, at which Exchange, and whether it is the buying or selling price. However, "the closing price of 10,000 bushels of wheat on June 30, 1980 at the Chicago Commodity Exchange" is a well-defined quantity.

11

- Describe the quantity on a scale that is meaningful to the subject. For example, if the uncertain quantity refers to a quantity of oil, the subject may think in terms of gallons, barrels, or tank cars. The wrong choice of scale may cause the subject to spend more effort on fitting his answers to the scale than on evaluating his uncertainty. It is important, therefore, to choose a unit with which the subject is comfortable. After the encoding, the scale can be changed to fit the analysis. As a rule, let the subject choose the scale. It may sometimes be easier to express an uncertain quantity as a fraction of another quantity or as the excess over another quantity rather than in absolute terms.

2.4 Joint Distributions

Most decision problems will have more than one state variable for which the uncertainty should be encoded. Part of the modeling effort is to determine the dependencies among the variables and perhaps to redefine the variables to make new variables that are independent.

Two variables are said to be independent if the probability distribution for one of the variables does not depend on the revealed value of the other variable. For example, the unit manufacturing cost and the average annual mileage for a new truck would be considered independent if the uncertainty in mileage did not depend on the level of the unit manufacturing cost. As a contrast, the engine reliability and the average monthly maintenance cost would probably not be considered independent because a high maintenance cost might be more likely when the engine reliability is low than when it is high.

If the variables are judged to be independent, their uncertainties can be encoded separately. Otherwise, the uncertainty in each variable has to be encoded separately for each possible outcome of the variables on which it depends. Dependence increases the labor of the analysis. A discrete example demonstrates this; if each of three variables is modeled to have only three outcomes, thirteen ($1 + 3 + 3 \times 3$) distributions have to be encoded in the case of dependence as compared with only three in the case of independence. The effect of dependence increases rapidly when the number of possible outcomes increases. It is

12

obvious that one should try to model in such a manner that variables are independent. The following paragraphs present examples of how such modeling can be achieved.

A set of variables may be considered dependent because each of them depends in some way on the same underlying variable. However, they may be independent for each value of the underlying variable. Total annual fuel cost and total annual maintenance cost may be considered dependent because both are related to the total annual mileage of trucks; a high total fuel cost is indicative of a high total maintenance cost. One way to restructure the model is to use the following three variables that may be considered independent: total annual mileage, average fuel consumption, and average maintenance cost (per mile). Given these variables, it is easy to calculate total annual fuel cost and total annual maintenance cost.

For a second example, consider the demands, X1 and X2, for two similar products in a product line in a given period. X1 and X2 are dependent because the two products are competing in the same market. At the same time, it is possible that the first product's share of the market, Y1, is judgmentally independent of the total market, Y2, for the two products. We then encode the uncertainties for the two new quantities Y1 and Y2 and use the following obvious relationships to get back to the original demands:

$$X1 = Y1 \times Y2 , \quad \text{and} \quad X2 = (1 - Y1) \times Y2 .$$

13

3 BIASES IN JUDGMENT OF UNCERTAINTY

3.1 Relevance for Probability Encoding

Probability encoding is defined as the process of translating a person's judgment about an uncertain quantity into a probability distribution. There is an abundance of experience available, from practical applications of encoding as well as from experiments, showing that it provides useful information but that it is often biased. The purpose of this chapter is to describe some important forms of biases and situations in which they may appear. The awareness of potential biases may help an analyst detect and correct such biases. Some corrective procedures are presented as part of the interview process in Chapter 5; a more detailed discussion of biases and corrective procedures is included in Appendix A.

People seem to perceive and assess uncertainty in a manner similar to the way they perceive and assess distance. They use intuitive assessment procedures that are often based on cues of limited reliability and validity. Generally, these procedures, or modes of judgment, produce reasonable answers. For example, a captain of a ship can generally estimate distance accurately enough to avoid accidents, and a business executive can generally evaluate uncertainties well enough to make his enterprise profitable. On the other hand, overreliance on certain modes of judgment may lead to answers that are systematically biased, sometimes with severe consequences.

People consistently overestimate the distance of a remote object when visibility is poor and underestimate the distance when the sky is clear. In other words, they exhibit a regular systematic bias. This is because they rely on the haziness of an object as a cue to its distance. This cue has some validity because more distant objects are usually seen through more haze. At the same time, this mode of judgment may lead to predictable errors.

15

Three features of this example are worth noting. First, people generally are not aware of the cues on which their judgments are based. Few people consciously realize that they use haze to judge distances, although research shows that this applies virtually to everybody. Second, it is difficult to control the cues people use; the object seen through haze looks more distant, even when we know why. Third, people can be made aware of the bias and can make a conscious attempt to control its effects, as an aircraft pilot does when flying on a hazy day.

These same characteristics pertain to the assessment of uncertain quantities. Here too, people are usually not aware of the bases of their judgments, which often introduce systematic biases. Likewise, although modifying impressions and intuitions is exceedingly difficult, it is possible to learn to recognize the conditions under which such impressions are likely to lead to biases. Appropriate corrective procedures may then be applied.

People often deal with uncertainty by avoiding it or at least by being very cautious. With better ways to understand uncertainty, the decision maker would know when to exercise caution and when less conservative behavior could improve his performance. This is similar to the airplane pilot who, with proper instrumentation, can fly safely through fog and who, without it, does not try.

3.2 Biases in Probability Encoding

For the purposes of this section the subject is assumed to have an underlying knowledge regarding the quantity under investigation. This knowledge may be changed through receiving new information. The task of the analyst is to elicit from the subject a probability distribution that describes the underlying knowledge. Conscious or subconscious discrepancies between the subject's responses and an accurate description of his underlying knowledge are termed biases.

Biases may take many forms. One is a shift of the whole distribution upward or downward relative to the basic judgment; this is

16

called _displacement bias_. A change in the shape of the distribution
compared with the underlying judgment is called _variability bias._
Discrepancies may be a mixture of both kinds of bias. Variability bias
frequently takes the form of a _central bias_, which means that the
distribution is narrower than is justified by the subject's actual state
of information. A central bias is the effect of overconfidence; the
subject is expressing greater confidence than is justified. Of course,
it is difficult to say that a particular probability distribution is too
narrow. But there is overwhelming experience that the revealed value of
a quantity is found much more often in the tails of the distribution (or
even beyond the endpoints) than should be the case. Biases are
illustrated in Figure 2 in the form of three density functions, where A
represents the underlying judgment and distributions B and C represent
the effects of central bias and displacement bias.
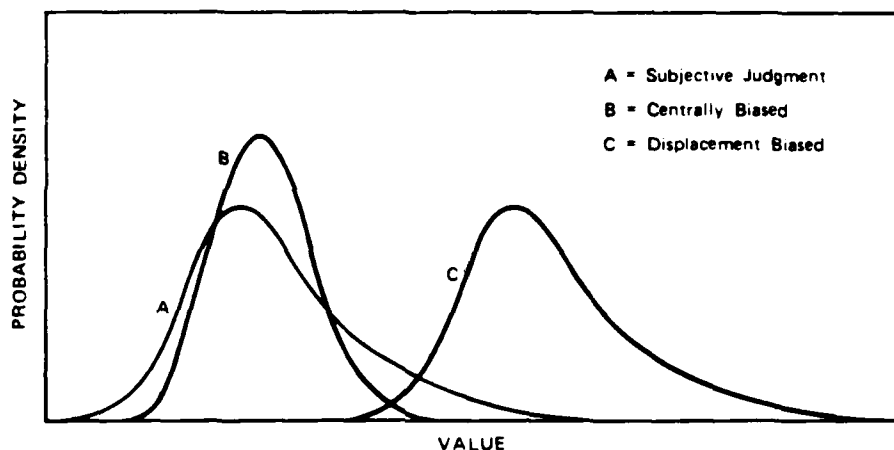
FIGURE 2   EXAMPLES OF VARIABILITY AND DISPLACEMENT BIASES

The sources of biases can be classified as motivational or
cognitive. _Motivational biases_ are either conscious or subconscious
adjustments in the subject's responses motivated by his perceived system
of personal rewards for various responses. He may want to influence the
decision in his favor by giving a particular set of responses. Or he
may want to bias his response because he believes that his performance

17

will be evaluated by the relationship between this response and the outcome. For example, a project manager may give a high estimate of the performance of a new project to ensure that the project will continue to receive adequate funding. On the other hand, a sales manager may give a low prediction of sales because he thinks he will look better if the actual sales exceed his forecast, which may be viewed as a commitment. Finally, the subject may suppress the full range of uncertainty that he actually believes to be present because he feels that someone in his position is expected to know what will happen in his area of expertise.

Even when a subject is honest--in the sense that he has no motivational bias--he may still have cognitive bi ses. Cognitive biases are either conscious or subconscious adjustments in the subject's responses systematically introduced by the way the subject is intellectually processing his perceptions. For example, a response may be biased toward the most recent piece of information simply because that information is the easiest to recall. Cognitive biases, therefore, depend on the modes of judgment used by the subject. They will be discussed further in the next section.

3.3 Different Kinds of Biases

An important responsibility of the interviewe is to be prepared for biases and try to adapt the interview to minimize them. In this section, we will describe different kinds of biases in intuitive judgment and give examples of how they might operate. These biases are further discussed in Appendix A.

3.3.1 Specific and General Information

A person who is making an assessment of an uncertain quantity is likely to consider information about similar quantities as well as information that pertains specifically to the quantity at hand. We call the two kinds of information general and specific information.

For example, consider the annual maintenance cost for a new type of aircraft. Specific information would refer to the detailed design of

18

the aircraft, the materials used for various parts, and the potential uses of the aircraft. General information, on the other hand, would concern maintenance experience with similar aircraft but possibly of other sizes, designs, or applications. It seems reasonable to make use of the broader experience available as general information and not to start from scratch for every new assessment.

It should be mentioned, however, that there are many quantities for which little, if any, relevant general information is available. In particular, this is the case with long-term forecasts, such as for the exchange rate of the German mark to the U.S. dollar in 1990 or of the price and supply of crude oil after the year 2000. In such cases, one will have to base the estimate on specific information alone.

3.3.2 Nonregressive Estimation

People often base an estimate of an uncertain quantity on a few characteristics of the quantity or on their impression of the quantity. Many times this impression is made on the basis of specific information about the quantity. There is substantial evidence that people tend to rely primarily on specific information, even when it is limited and unreliable (and even when it is almost worthless), and that they give insufficient weight to general information. In fact, the tendency to neglect general information, or at least to give it too little weight, may be the most important error in intuitive judgment. The resulting bias is called nonregressive estimation. The effect of the bias is to shift the whole distribution for the quantity upward or downward; thus, it is a displacement bias.

Returning to the example of estimating the maintenance cost of a new aircraft, it may be easy to consider only the design and to evaluate different subsystems so as to determine annual maintenance costs. The conclusion may be that the aircraft represents an improvement in terms of maintenance over earlier types. However, it would be wise to recall the experience of other, but similar, aircraft and the relation between experienced maintenance costs and the maintenance costs forecast at the

19

time of introduction. More likely than not, this would justify an estimate somewhere between the original estimate and one representing the average experience with similar aircraft. This is called regressive estimation, or estimation toward the mean.

As another example, consider the sales of a newly developed machine. People in product development and market research state that the machine should be a success because of its low price and of a number of new features not available on similar machines. The overall impression is that the machine should be a high-volume product. If an estimate were made of the sales development, it would likely be a high number. However, this estimate fails to consider past experience with similar products, which constitutes general information. Looking back on the history of other products that were expected to be best-sellers, one would probably find that many such products failed to meet expectations. That is, the sales of new products are highly uncertain and are therefore difficult to predict. The best estimate of the sales of the machine should then fall somewhere between the value that matched the impression and the average value for sales of similar products.

The specific information often contains an element of time. For example, the general information may refer to the accumulated experience regarding a foreign country's buildup of forces or a competitor's market strategy, whereas the specific information may refer to a particular recent piece of information--e.g., an intelligence report on the foreign country or the competitor. Again, the tendency is to rely too strongly on this recent information, even though it may be unreliable, and discard past experience.

Nonregressive estimation occurs easily in repetitive situations. For example, the sales forecast for the coming year for a product that did exceptionally well last year is likely to be very high, whereas the best estimate would be less extreme considering the uncertainty in year-to-year variations.

When the subject is using his specific impression as the basis for his estimate, it is important to bring up relevant general information

20

and to compare the estimation task with the experience with similar quantities in the past. The estimate should fall somewhere between the initial impression and the average of the class; this will depend on the subject's ability to predict differences among members of the class on the basis of the specific information.

The predictability is really made up of two parts. One part relates to the uncertainty that is inherent in the estimation task even when the information is well specified. For example, week-to-week variations of stock prices are notoriously difficult to predict, whereas variations in tide levels are fairly easy to predict. The other part relates to the quality of the specific information--i.e., the validity (the relevance of the specific information for the estimation task) and the reliability (the uncertainty in the specific information given the outcome of the uncertain quantity). For example, a rumor regarding political activity in a foreign country may not be valid information; an intelligence report regarding enemy or competitor action may be valid but not reliable in that it is not highly correlated with actual action.

If the predictability is low, one should resort to the average of the class. The estimate should only match the impression from the specific information in the case of perfect predictability--i.e., when there is no uncertainty regarding the outcome of the quantity. This concept of regression toward the mean can generally be explained to the subject in a qualitative way: "When predictability is limited, things are rarely as good as one hopes nor as bad as one fears." If the subject accepts the argument, then he should be willing to modify his intuitive estimate.

In general, the cure for this bias is to ask the subject to consider the situation prospectively from the past: How would he have made his estimate before receiving specific information, thus basing his estimate solely on general information about similar quantities? How would he adjust this prior opinion for the specific information, considering the reliability and validity of the latter?

21

As mentioned above, the specific information frequently involves a time element, i.e., basic experience is modified by a recent piece of information. For example, a company had to decide whether or not to introduce a new product that was considered to have a high demand potential. The product was test-marketed, and there was a slightly unfavorable outcome; the revised assessment of the market said there was a low demand. This revision was made despite past experience with similar market tests that had been less than accurate in predicting the final market size and in contrast to the strong prior judgment indicating a high demand. This is a case of focusing on specific information and of ignoring general information, which perhaps should carry the main weight in the probability assignment.

In this example, the bias can be counteracted by considering the market forecast prospectively from the past. The probabilities, prior to the test result, for various levels of demand are encoded first. Then, the probability distribution for the test result conditional on the demand is encoded. A simple application of probability calculus then provides the probability distribution for demand given the outcome of the market test.

3.3.3 Availability

Probability assignments are based on information that the subject recalls or visualizes. The probability of a breakdown in a production process may be assigned by recalling past breakdowns. The probability that a politician will not be reelected to his office may be evaluated by considering various ways in which he may lose support. It is clear that the information used to form a probability assignment represents only a part of the subject's past experience; it is that part that is the easiest to retrieve from memory.

Availability refers to the observation that the easier it is to think of occurrences of an event, the more likely the event will be considered. Thus, availability is concerned with the ease with which relevant information is recalled or visualized. It is easy to recall

22

information that made a strong impression at the time it was first presented. Recent information is more available than old information and is often given too much weight. Some events may become overly available because of their potentially disastrous consequences (e.g., an accident with a nuclear reactor) and thus may be assigned probabilities that are too high. Information pertaining to the occurrence of an event may be received from different sources; if the information is originating from the same source, it is redundant but will become more available and lead to overly high probabilities. Certified information, e.g., a strategic plan, is more available than uncertified information.

Probabilities will generally be too high because of availability when information is easily recalled because it is recent, dramatic, redundant, certified, or salient. They will be too low when information is difficult to recall because it is old, undramatic, unique, uncertified, or minor.

Availability is essentially defined in the context of probability assignments for discrete events. Discrete events may be key uncertainties in many applications of decision analysis. Examples might be whether a new chemical process will work, whether there will be a competitive product on the market within 5 years, whether there will be a government ban on a certain production process, and whether there will be a major war in the Middle East before 1985. However, the availability of information may also lead to a displacement bias for continuous probability distributions.

The cure for availability is to ask the subject to consider the situation retrospectively from the future: Assuming that the event did not occur, or that there was an outcome quite different from the one discussed or estimated, how could that have happened? Could the subject generate a scenario to explain the outcome? This procedure will make other information available to the subject and thereby make the original probability assignments less extreme.

3.3.4 Anchoring Biases

The most readily available piece of information often forms an initial basis for formulating an estimate; the estimate then represents an adjustment from this basis. For example, the current business plan is often used as an available starting point. Likewise, when predicting this year's sales, the subject may use last year's sales as a starting point.

There is nothing wrong in using available information; on the contrary, it would be foolish not to use it. However, the subject's adjustment from such a starting point is often insufficient, and we say that the starting point is serving as an anchor. Thus, anchoring can occur when some information has become overly available at the beginning of the interview. Anchoring results from a failure to generate other information that also might be relevant for the estimation of the quantity at hand. The effect is primarily one of displacement bias.

Even seemingly irrelevant information may serve as an anchor, particularly when the subject is feeling very uncertain about a quantity. The way a set of questions is asked may unconsciously be used as information. The same is true with any value mentioned to the expert. Even defining the scale for the uncertain quantity makes information available. If the interviewer chooses an unnatural scale, this may easily lead to a displacement bias. For instance, using a scale in millions of dollars when the revealed value is likely to be in billions of dollars usually leads to underestimation. It is therefore important to let the subject choose the scale.

Anchoring will also produce a central bias. Initial estimates and easily available information may serve as anchors for subsequent responses. These responses are made as adjustments from the anchor, and in general these adjustments are not sufficient. The initial response in an interview often serves as a basis for later responses, especially if the first question concerns a likely value for the uncertain quantity. Anchoring results from a failure to process information about other points on the distribution independently from the anchoring point.

24

Experimental results show, for example, that subjects seem to produce a central bias when they are asked first for the median for an uncertain quantity and then for the quartiles. Subsequent responses seem to be anchored on the first response, the median, which the subject usually views as the best single-number forecast.

Anchoring may occur even though the interview does not begin with a likely value. The subject may elect to begin to think about a central value before responding to actual questions.

The anchoring effect is prone to appear in almost every encoding session. It is therefore important to be able to counteract it. This can be done in different ways. One way is simply to be careful in the interview not to supply numbers or information that may serve as anchors. Another means is to reduce the effect of what seems to be an anchor by either supplying other anchors that would pull the estimate in the opposite direction or by making the subject generate such information. The general cure is to ask the subject to consider the situation retrospectively from the future: Assuming that an extreme outcome (high or low) occurred, could he generate a scenario to explain the outcome?

So far, we have discussed anchoring only on a value for an uncertain quantity. Anchoring can also occur when a subject is assigning a probability to a discrete event. The natural anchor is to assign equal probabilities to all events. This is most notable for the dichotomous case--i.e., a case with only two outcomes: either an event occurs or it does not. "Fifty-fifty" then serves as a natural starting point, and the final probability assignment represents an adjustment from that point that often is not sufficient. The cure is to ask the subject to generate scenarios that would lead to each of the two outcomes and thereby differentiate between them. A visual display of the two probabilities--e.g., with a probability wheel (see Section 4.2.1)--will also be useful in showing the relationship between the two probabilities.

3.3.5 Unstated Assumptions

A subject's responses are typically conditional on various unstated assumptions. Consequently, the resulting probability distribution does not properly reflect his total uncertainty. For example, the subject may not have considered such possibilities as future price controls, major strikes, currency devaluation, war, and so on, when expressing his judgment because he assumes that he is not responsible for considering such events. One result is that he may be less surprised than might be expected when the revealed value of an uncertain quantity falls outside the range of his distribution. He justifies this because of a drastic change in some condition that he did not feel responsible for incorporating into his judgment.

Although the subject cannot be held responsible for taking into account all possible eventualities that may affect the quantity he is assessing, it is his (and the interviewer's) responsibility to state the assumptions he is making about his own limits of responsiblity. Once identified, they can be built into the model, and an appropriate expert (who may or may not be the current subject) can assign their probabilities.

The change of the time perspective to look at the situation retrospectively from the future, as described above, may reveal some unstated assumptions. That is, the subject may explain an extreme outcome by a condition that he had not included in his original judgment.

3.3.6 Coherence

People sometimes appear to assign probabilities to an event based on the ease with which they can fabricate a plausible scenario that would lead to the occurrence of the event. The event is considered unlikely if no reasonable scenario can be found; it is judged likely if many scenarios can be composed that could make the event occur, or if one scenario is particularly coherent. The credibility of a scenario to a subject seems to depend more on the coherence with which its author

26

has spun the tale than on its intrinsically 'logical' probability of occurrence. The coherence of the scenario even tends to overrule the fact that the scenario may be based on a sequence of enabling events, some of which may not be very likely if looked at separately.

For example, the probability assigned to the event that sales would exceed a high volume may depend on how well market researchers have put together scenarios that would lead to that volume; these could be scenarios on what markets might be penetrated and what the penetration rate might be with a reasonable marketing effort. Courtroom arguments are another case in which evaluation of credibility is often based on the coherence of the sequence of evidence as presented by the prosecution or the defense. It is thus important that the discussion of possible outcomes for an uncertain quantity be well-balanced because the relative coherence of various arguments can have a strong effect on the probability assignments.

The cure against the coherence bias is to ask the subject to generate scenarios that would lead to other outcomes. This can be done, as before, by looking at the situation retrospectively from the future.

3.3.7 Planning Biases

This group of biases often occurs with estimates of quantities that are building on a complex set of events or quantities. They typically occur in planning situations and may concern the time to complete a project, the cost of developing a new product, the market share of a product in a new market, etc. There are generally many things that can go wrong, which taken together would lead to an increase in time or cost or to a reduced market share. But the joint impact of these factors is often underestimated because each of them is so unlikely to occur.

These biases may be reduced if the subject is asked to list all factors that may have a negative impact. He may then become more aware that the probability of something going wrong no longer is negligible.

27

3.3.8 Time Biases

People often neglect the time aspect of an event. For example, when assigning a probability to the event that there will be a major war in the Middle East, they may give similar assignments whether the event is defined for the next month or for the next year. This leads to an overestimation for events defined for a short time period and to an underestimation for events defined for a long time period.

If the interviewer suspects a time bias, he may ask the subject to assign probabilities to the event occurring in time periods of different lengths. If the probability assignments do not reflect the time element, this can be discussed with the subject, who can then be asked to revise his original probability assignment.

4   ENCODING METHODOLOGY

This chapter describes the most important probability encoding techniques of which we are aware. These techniques are classified according to the encoding method and the response mode used. The interview process is further explained in the next chapter.

4.1   Encoding Methods and Response Modes

Most encoding methods are based on questions for which the answers can be represented as points on a cumulative distribution function. The different encoding methods vary according to whether they ask a subject to assign probabilities (P), values (V), or both. The three basic types of encoding methods are listed below:

- P--methods require the subject to respond by specifying points on the probability scale while the values remain fixed.

- V--methods require the subject to respond by specifying points on the value scale while the probabilities remain fixed.

- PV--methods ask questions that must be answered on both scales jointly; the subject essentially describes points on the cumulative distribution.

Any encoding procedure consists of a set of questions that the subject responds to either directly by providing numbers or indirectly by choosing between simple alternatives or bets. In the direct response mode, the subject is asked questions that require numbers as answers. Depending on the method being used, the answers will be given in the form of either values or probabilities.

In the indirect response mode, the subject is asked to choose between two or more bets (or alternatives). The bets are adjusted until he is indifferent in choosing between them. This indifference can then be translated into a probability or value assignment. When an external

reference process is used, one bet is defined with respect to the uncertain quantity and the other with respect to a similar reference event. That is, each bet is made on the occurrence of an event that is related to either the uncertain quantity or the reference process. The reference process can be of a symmetrical type (e.g., a wheel of fortune or a deck of cards) or refer to an event for which no symmetry can be perceived (e.g., the crash of a regular air carrier between San Francisco and New York).

The subject can also be asked to choose among events defined on the value scale for the uncertain quantity, where each event represents a set of possible outcomes for the uncertain quantity (e.g., the event of export sales of weapon systems being less than or equal to 2,000 units or being greater than 2,000 units). This type of response mode uses internal events for comparison.

Each probability encoding technique can be classified according to the encoding method and response mode used. The techniques that we have found most useful are listed in Table 1 and described below.

4.2 Indirect Response Techniques

4.2.1 Probability Wheel

The probability wheel is one of the most useful tools that we have discovered for eliciting unbiased responses from subjects. The wheel is a disk with two adjustable sectors, one blue and the other orange, with a fixed pointer in the center (see Figure 3). When spun, the disk will finally stop with the pointer either in the blue or the orange sector. A simple adjustment changes the relative size of the two sectors and thereby also the probabilities of the pointer indicating either sector when the disk stops spinning. The subject is asked which of two events he considers more likely--the event relating to the uncertain quantity (for example, the event that next year's budget will not exceed X units), or the event that the pointer ends up in the orange sector. The amount of orange in the wheel is then varied until the subject finds the

Table 1

CLASSIFICATION OF PROBABILITY ENCODING TECHNIQUES

Encoding Method	Response Mode		
	Indirect		Direct
	External Reference Events	Internal Events	
Probability (value fixed)	Probability wheel	Relative likelihoods	Cumulative probability
Value (probability fixed)	Probability wheel Fixed probability events	Interval technique	Fractiles
Probability-Value (neither fixed)	--	--	Drawing graph; Verbal encoding

FIGURE 3   A PROBABILITY WHEEL

31

two events equally likely. The relative amount of orange is then assigned as the probability of the event.

Sometimes it helps to rephrase the questions in terms of choices among bets. For example, the subject is offered a choice between two propositions: he can win some amount, say $100, if the budget does not exceed X units; he can win the same amount if the pointer ends up in the orange sector. This is an example of a P-method in which the event (the value) is fixed and the subject is asked to assess the probability. The probability wheel can also be used as a V-method; however, this is generally less effective, primarily because the P-method asks the subject to evaluate the uncertainty for a single event, whereas the V-method requires that the subject evaluate a sequence of events until he finds one that he considers equally likely as the reference event.

One advantage of the probability wheel is that it simultaneously displays the probability that an event does not occur and that the event will occur. Thus, the subject is always aware that there is a trade-off between the probabilities for an event and for its complement. Another advantage of the probability wheel is that the probability can be varied continuously from 0 to 1. However, because it is difficult for a subject to discriminate between the sizes of very small sectors, the wheel is most useful for evaluating probabilities in the range from 0.1 to 0.9.

There are other tools that help display the probability of an event (and its complement) and that can be used instead of the probability wheel. One alternative is a horizontal bar with a movable marker that can be set to define two events--one to the left and one to the right of the marker. Another alternative is to ask the subject to visualize an urn with, say, 1,000 balls of two colors. A ball is supposedly drawn at random from the urn, and the reference event in this case is that the ball drawn will be orange. The composition of the urn can then be varied until it reflects the probability of the event in question. We prefer to use the probability wheel because subjects find it easier to

visualize the chance process by looking at the wheel than by using the bar or visualizing the urn.

4.2.2 Fixed Probability Events

Other less tangible reference events may be useful in the encoding process, particularly when reference has to be made to low-probability events. For example, the event of tossing ten 'heads' in a row with an unbiased coin has a probability close to 1/1000. Some subjects can identify easily with events relative to poker hands. For example, a royal flush has a probability of roughly 1/65,000. Typically, reference processes such as these concern events with fixed probabilities and therefore basically work as V-methods, wherein the subject is asked to assess values that correspond to fixed probabilities. They could be used as reference events for P-methods if there were a large enough set of them: one would then go through the list of fixed probability events until indifference was reached.

4.2.3 Interval Technique

The interval technique is an example of an internal events response mode and is a V-method. To begin this technique, an interval is split into two parts. The subject is then asked to choose which part he would prefer to bet on, or which of the two parts he considers most likely. The dividing point is changed to reduce the size of the part considered most likely by the subject (thereby increasing the size of the other part), and the subject is asked to choose between the two new parts. The position of the dividing point is adjusted until the subject is indifferent between the two parts. These subintervals are then assigned equal probabilities. Beginning with an interval covering all possible outcomes and then splitting into two subintervals first gives the median, then the quartiles, and so on. It is usually not meaningful to continue the interval technique after the quartiles have been obtained because each question depends on earlier responses and errors are thus compounded. Subjects tend to arrive at their quartile estimates by

33

adjustments from the median. These adjustments are generally insufficient and result in distributions that are too narrow.

We have found the analogy with a roulette game useful when encoding the median and quartiles: The range of possible outcomes for the uncertain quantity is divided into a number of intervals, and the subject is given the opportunity to bet on one of these intervals. He is supposed to have a single chip and will win a hypothetical prize (say, $1,000) if he places the chip on the interval that actually occurs. The layout of the game is shown in the diagram below:

Outcome of Betting	Definition of Bets					
Median	Below		230		Above	
Quartiles	A	200	B	C	270	D
Consistency Check	Outside		Inside			Outside

The object of the session is to fill in the three numbers in the diagram. The subject is first asked to bet on the outcome being below or above a certain number. That number is then changed until indifference is reached; it is then the median (it becomes 230 in the illustrative diagram). Next, the subject is given the opportunity to bet on either Interval A or Interval B (where the upper limit of B is the median). The dividing point is changed until indifference is reached; it is then the lower quartile (200 in the diagram). The upper quartile is found in a similar way by bets on Interval C or D. As a final check, the subject is asked whether he would prefer to bet on "inside" (Intervals B and C) or on "outside" (Intervals A and D). If he is not indifferent, then this inconsistency should be explained to the subject and the set of responses should be reviewed and possibly changed.

34

The interval technique can also be based on splitting the interval into three parts. (Using the technique for more than three parts is not recommended.) The subject is then asked to rank the three parts from the part he considers most likely to the one he considers the least likely. The two dividing points are then changed to reduce the size of the most likely subinterval and to increase the size of the least likely subinterval (the size of the third part should remain about the same as before), and the subject is again asked to rank the three subintervals. When two subintervals are considered equally likely (and each is, say, less likely than the third part), then the dividing points are changed to increase simultaneously the sizes of the two subintervals and to reduce the size of the third subinterval. The procedure is continued until indifference among the three parts has been reached.

4.2.4 Relative Likelihoods

A P-method that uses the internal events response mode asks the subject to assign relative likelihoods (or odds) to two well-defined events. For example, the subject may first be asked whether he considers next year's export sales of weapon systems more likely to be above or below 5,000 units. The next question is then: How many times more likely is it? This method is used primarily for uncertain quantities that have only a few possible outcomes.

4.3 Direct Response Techniques

4.3.1 Cumulative Probability and Fractiles

In the direct response mode, the subject can be asked either to assign the cumulative probability at a given value (e.g., what is the probability that next year's export sales of weapon systems will be less than or equal to 3,000 units?), or to assign the value corresponding to a probability (e.g., what is the level of sales that corresponds to a 10% probability?). The probability response can be expressed either as an absolute number (0.20), as a percentage (20%), or as a fraction (1 in

5, or 2 in 10). This last way is particularly useful for small probabilities because subjects usually can discriminate more easily between 1 in 100 and 1 in 1,000 than between the absolute numbers 0.01 and 0.001. Expressing a probability in fractional form is closely related to expressing it in terms of odds or ratios, particularly for probabilities close to 0. The probability $1/n$ is equivalent to the odds $1:(n-1)$ (generally, odds of $m:n$ in favor of an event correspond to the probability $m/(m + n)$ for the same event), and the two numbers are close enough for most practical purposes for values of n greater than 100. An encoding session often will include a few questions relating to low-probability events; we expect more answers to be in the form of probabilities expressed in fractional form than in odds form. However, we will use the term 'odds questions' to cover both cases.

The form chosen to express probability should be the one that is most familiar to the subject.

4.3.2 Graphs

Graphing uses a direct response mode and requires a subject to provide joint probability and value assignments, thereby making it a PV-method. It requires that the subject either draw a density function or a cumulative distribution or state a series of pairs of numbers (value and probability). Another approach is to show the subject a series of density functions and then ask him to choose the one that corresponds most closely to his judgment. The density functions can be generated easily by taking a family of distributions (e.g., beta distributions) and varying the parameters. With the help of CRT displays, this has been done in some psychological experiments in which the subject has used two levers to change the parameters and thereby vary the displayed density function.

4.3.3 Verbal Encoding

Verbal encoding uses verbal descriptors to characterize events in the first phase of the encoding process. The descriptors used are

36

those, such as 'high,' 'medium,' and 'low' procurement cost, to which the subject is accustomed. Quantitative interpretation of the descriptors is then encoded in a second phase. This method could be of particular use in dealing with quantities that have no ordinal value scale. Like the graphing technique, verbal encoding is a PV-method that requires the subject to provide joint probability and value assignments.

4.4 Applicability of the Various Techniques

Subjects seem to fall into two general categories: those who feel capable of (and often prefer) giving direct numerical probability assignments and those who experience difficulty in making such judgments. Most people seem to fall into the second category. Furthermore, many individuals who prefer direct numerical responses are later found to have little confidence in their initial numerical responses. For this reason, the indirect response mode is generally the better way to begin encoding. Of these techniques, the probability wheel is the one most subjects find easiest to use. Later, the *interviewer can shift to the direct response mode if he believes that the subject will give the same responses as with the wheel (or if the subject has stated that he would rather not use the wheel).* However, we prefer to use the wheel as long as possible to keep the subject aware of the fact that assigning a probability to an event automatically means assigning the complementary probability to the complementary event; this is explicitly illustrated with the wheel.

The interval technique, which asks a subject to generate the median, quartiles, and sometimes tertiles, is especially useful for arriving at a meaningful assessment of the median. However, it is generally unwise to begin the encoding process by eliciting the median because that value tends to serve as an anchor for subsequent responses. Instead, the preference is to use the interval technique as a consistency check after other techniques have been used. Similarly, relative likelihood questions are usually used for verification of *earlier responses.*

37

Subjects are seldom able to express their uncertainty in terms of a density function, a cumulative distribution, or moments of a distribution. Therefore, it is usually not meaningful to try eliciting a distribution or its moments directly. There are, for example, procedures that ask the subject for the parameters of a special distribution--for example, the mean and standard deviation of a normal distribution or a beta distribution. Our experience indicates that subjects will give such parameters, but that usually they do not understand the full implications. We believe that the choice of special distributions is a modeling consideration and should not normally be made part of the encoding process. However, we do find that graphical displays of distributions drawn from indirect responses can provide useful feedback to the subject.

5 THE INTERVIEW PROCESS

Any procedure used to help in making a decision that depends on the judgment and knowledge of different people is vulnerable to *their biases and prejudices*. Decision analysis is equally vulnerable, but if the encoding process is carefully carried out, these biases may often be discovered or avoided. A poorly executed probability encoding session will lead to a loss of credibility for any analysis based on this judgment. Furthermore, the subject is allocating time and effort to the encoding process, and *thus the best incentive for him* to make this allocation is an efficient encoding procedure. If he is asked questions that are either difficult to understand or seemingly irrelevant, obtaining his cooperation at some later time may be more difficult.

A good interview process need not be elaborate, but should ensure that the proper questions are asked. While the structure of the interview process is still evolving, the following approach has been found to be effective. The process is divided into five phases.

- Motivating--Rapport with the subject is established and possible motivational biases are explored.

- Structuring--The uncertain quantity and the structure underlying it are precisely and unambiguously defined.

- *Conditioning*--The effect of some potential biases is reduced and the subject is conditioned to think fundamentally about his judgment.

- Encoding--The subject's judgment is quantified in probabilistic terms.

- Verifying--The responses obtained in the encoding are checked for consistency.

These five phases are discussed in the following sections. Appendix B presents the interview process in more detail, and the individual steps are illustrated by sample questions and answers.

39

5.1 Motivating

This phase of the interview process has two purposes. The first purpose is to introduce the subject to the encoding task. This may entail an explanation of the importance and purpose of the probability encoding session and a discussion of the difference between deterministic (single number) and probabilistic (probability distribution) estimates.

The second purpose is to explore whether any motivational biases might operate. The interviewer and the subject should discuss openly any payoffs that may be associated with the probability assignment, and the ramifications of possible misuses of the information. The subject may be aware of misuses of single-number predictions--e.g., that they often are interpreted as firm projections or commitments. It should be pointed out that no commitment is inherent in a probability distribution and that the only aim of the encoding process is to develop a probability distribution that represents as clearly as possible the complete judgment of the subject.

If the subject is involved in some way with the uncertain quantity (e.g., the product manager is being asked about the sales potential), the discussion should be directed to factors that may affect the outcome of the quantity but that are outside the subject's control (e.g., competitive action). It should be easier for the subject to provide unbiased judgment if he understands that he cannot be held responsible for every aspect of the outcome.

During the discussion, the subject is likely to reveal some reactions indicating biases that might be expected later in the encoding process. For example, the subject may display some caution against overestimation and may therefore be somewhat biased toward underestimation. This knowledge will later influence the choice of encoding method.

5.2 Structuring

The purpose of this phase in the interview process is to define and clearly structure the uncertain quantity. The quantity is assumed to be important to the decision. It should be defined as an unambiguous state variable. The definition should pass the clairvoyant test: that is, a clairvoyant should be able to specify the outcome without asking additional questions for clarification. The structure should be expanded as necessary so that the subject does not have to model the problem further before making each judgment. There may be different ways of breaking down the structure; if so, the subject should be *included in making the choice of breakdown.* It is also important to let the subject choose a scale that is *meaningful to him.* This includes the question whether the quantity should be evaluated in *absolute terms or be expressed in relation to some other quantity* (e.g., as the increase over the corresponding value last year).

The subject should be required to think the problem through carefully before the actual encoding phase begins. He should decide what background information might be relevant (or irrelevant) to the problem. Otherwise, only the readily available information will be used initially, and new information may surface later in the session and invalidate all previous answers. The interviewer should probe any areas that seem unclear to either party. This phase will vary greatly, depending on the decision problem, the uncertain quantity, and the subject.

5.3 Conditioning

The purpose of this phase of the interview is to draw out the subject's knowledge relating to the uncertain quantity. This procedure serves to give him a conscious basis for making probability judgments and to counteract encoding biases that he might otherwise exhibit.

5.3.1 Balancing General and Specific Information

The discussions in the first two phases of the interview will generally have revealed the information on which the subject is basing his judgment. Otherwise, the interviewer should now try to bring this up in the discussion. There will be situations in which all of the information is essentially either general or specific; in these cases, there will be no balancing problem, and the interview can move on to the steps described in Subsection 5.3.2.

In many cases, however, the subject will have both general and specific information about the uncertain quantity. The interviewer should then find out whether the estimate of the uncertain quantity will be based primarily on the specific information; if this seems to be the case, the interviewer should ascertain whether the subject assigns a high degree of predictability to the outcome of the quantity given the specific information. A bias is likely to occur when the predictability is not high, but the subject still bases his judgment on the specific information. The aim of the process is then to reach a reasonable *estimate of the uncertain quantity* by properly balancing general and specific information. This can be accomplished in several different ways, depending on how well defined the specific information is.

The most straightforward case is the one in which the specific *information is well defined.* This is the case, for instance, when the specific information relates to a new piece of information: the problem may be to estimate the performance of a new system after some information has been received from a field trial. The subject should then be asked to think back on what estimate he would have made prior to receiving the specific information. He is then asked to consciously consider the specific information, its validity and reliability, and what its impact really should be. Finally, he is asked for a revised estimate of the uncertain quantity that includes all of his knowledge, both general and specific.

In rare situations, the above calculations can be carried out formally as probability revisions because of the importance of the

uncertain quantity and the clear problem structure. The process then goes directly into the encoding phase (to be described in Section 5.4) and the subject's judgment prior to receiving the specific information is encoded in the form of a probability distribution. Next, the probability distribution of the specific information is encoded for all values of the uncertain quantity. Probability calculus then is applied to revise the probability distribution for the uncertain quantity on the basis of the distribution for the specific information. This way, any judgmental errors in combining different information are eliminated.

In other cases, one may obtain an estimate based on general information alone by asking the subject to identify, if possible, the uncertain quantity as a member of a reference class and to assess the average value and a range of variability for this reference class. The choice of a reference class is sometimes fairly direct; e.g., the cost overrun of a project can be related to the overruns of other projects by the same contractor or to overruns of similar projects. In other cases, the reference class can be obtained by a reformulation of the quantity. For example, the sales of a new product may be compared with the sales of similar products in terms of market share, and the unit manufacturing cost may be compared with that of other products in terms of the size of the cost relative to the budget value. Admittedly, however, it may not always be easy to define a relevant reference class (it may be particularly difficult with long-term forecasts); we will discuss this case below.

The subject is then asked to make an intuitive, top-of-the-head, estimate of the uncertain quantity and to assess the predictability--i.e., his ability to predict differences in values among quantities belonging to the reference class based on the kind of specific information that he has. In case the intuitive estimate is relatively extreme within the reference class and the predictability is assessed to be low or moderate, then the estimate should be regressed toward the average value of the reference class. The estimate should remain unchanged only when the predictability is high. In most cases,

explaining to the subject the basic principle of regression toward the mean will induce him to change his estimate. Alternatively, it is possible to ask him to revise the estimate given by the average for the class on the basis of the considered impact of the specific information.

When no reference class can be suggested, one can ask the subject to consider another expert who would make an estimate of the uncertain quantity based on general information about the quantity but without the specific information available to the subject. This 'less informed' estimate can then take the place of the average for the reference class and the process can continue.

5.3.2 Counteracting Anchors

When the subject's answers suggest that he may have reached his estimate by adjusting from a salient value or from a plan, there is reason to believe that the adjustment is insufficient--i.e., that the judgment is anchored. The interviewer can then point out that the adjustment probably is insufficient. He can also ask the subject to list other relevant values; these might then serve as other anchors that will help pull the estimate away from the first anchor.

The purpose of the next step is to bring to the surface more of the subject's knowledge on which he may base his judgment. On the basis of the previous discussion, the interviewer proposes some extreme values to the subject. The subject is then asked to regard the situation retrospectively from the future. That is, he is asked both to assume that he is told at some future time that such an extreme value had occurred and to describe a scenario that would explain this outcome. He should further be asked for the probability of outcomes outside of the extremes. When people are told an outcome has occurred, even hypothetically, they find it relatively easy to generate an explanation. This type of question serves to quickly surface more knowledge regarding the total range of possibilities and to bring up unstated assumptions the subject may have been making. Thus, this step may compensate for anchoring, availability, and unstated assumptions.

5.3.3 Compensating for Other Biases

In planning situations in which the subject has part of the responsibility for the outcome of the uncertain quantity, it will be helpful to ask the subject to make a list of external factors that might upset the plan. This will make it easier for him to realize the inherent uncertainty and to admit it in his judgment.

When a subject is assigning a probability to the occurrence or nonoccurrence of some event (for example, the probability that a system will or will not be successful in field tests), he may base his assignment on whether he can generate plausible scenarios leading to the occurrence of the event in question. Asking him to state the basis for his probability assignment may reveal that the coherence of such scenarios has been influencing his judgment. The interviewer may then want to generate more scenarios that would or would not lead to the occurrence of the event. For example, simply devising an equally coherent scenario that implies the opposite outcome might considerably change the subject's final probability assignment.

5.4 Encoding

The first three phases of the interview process have defined the quantity for which the uncertainty is to be encoded, the structure underlying the quantity, and the scale to be used for the quantity. They have eliminated or greatly reduced the effects of motivational biases, have brought forward the most important cognitive biases that may be operating, and have reduced their effects. The time has now come for the actual quantification of judgment in probabilistic terms. The procedures outlined for this phase of the interview process are provided as a guideline. They rest primarily on the use of the probability wheel as an encoding technique. Often a subject's responses will indicate a need to return to the tasks in the previous three phases. In particular, there may be a need for further structuring when the subject's responses and arguments indicate that they are based on different underlying assumptions.

45

Begin by using the probability wheel to encode the probability levels for a set of values. Take a value that you do not expect to be extreme (say, between the 0.2 and 0.8 fractiles) and encode the corresponding probability level. Do not choose the first value in such a way that may seem significant to the subject because this will cause him to anchor on that value. In particular, do not begin by asking for a likely value and then encoding the corresponding probability level.

Make the first few choices easy for the subject so that he will be comfortable with the task. This means, for example, that you should begin by making the orange sector on the probability wheel much smaller than what might actually correspond to the subject's probability. It is then easy for the subject to state which event is more likely and he becomes more comfortable with the procedure. Next, choose a sector that is much too large. After two easy choices, there is generally no problem to home in on the indifference point with a few more questions.

Continue to use the wheel for five to ten points, moving from one value to another without pattern. Ask for cumulative probability levels (the fixed event is defined as the quantity being less than or equal to a given value) or their complements (the fixed event is defined as the quantity being greater than a given value).

As you question the subject, plot each response as a point on a cumulative distribution and number the points sequentially. (It is also a good idea to use different plot marks for different encoding techniques.) An example is shown in Figure 4. This will point out any inconsistencies and will also indicate gaps in the distribution that need one or more additional points. Do not, however, show the plotted points to the subject at this stage in the process because he may try to conform to a smooth curve--i.e., he may try to make subsequent responses consistent with the plotted points.

Next use the interval technique to generate a value for the median and the quartiles. The interviewer must be aware that the interval technique often leads to quartiles that are too close to the central part of the distribution. When this seems to be the case, the
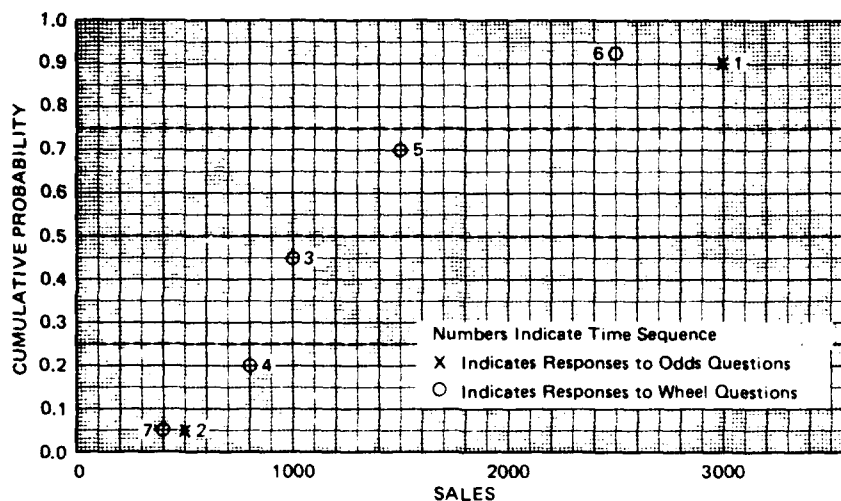
FIGURE 4   GRAPHICAL REPRESENTATION OF RESPONSES

interviewer will have to show the subject the discrepancies among responses generated by the wheel technique and by the interval technique and ultimately get the subject to bring them into consistency.

The order of the questions and of the different types of questions should be determined by the situation. The length of the encoding session depends on the ease with which the subject can answer the questions and on the convergence toward responses that are consistent with each other. Be alert to shifts in the subject's attention (for example, the shift of attention from the encoding process to the actual problem), changes in the subject's modeling of the situation, and the appearance of new information.

The encoding process is time-consuming, and it may be difficult to maintain a high motivation for the subject. It often helps to keep his interest, however, if he is shown some inconsistencies among his responses. Each response will lead to a point on a cumulative

47

distribution. The importance of the variable for the decision problem at hand determines the number of points to encode. After enough points have been encoded, a curve should be fitted to the points. An example is shown in Figure 5.
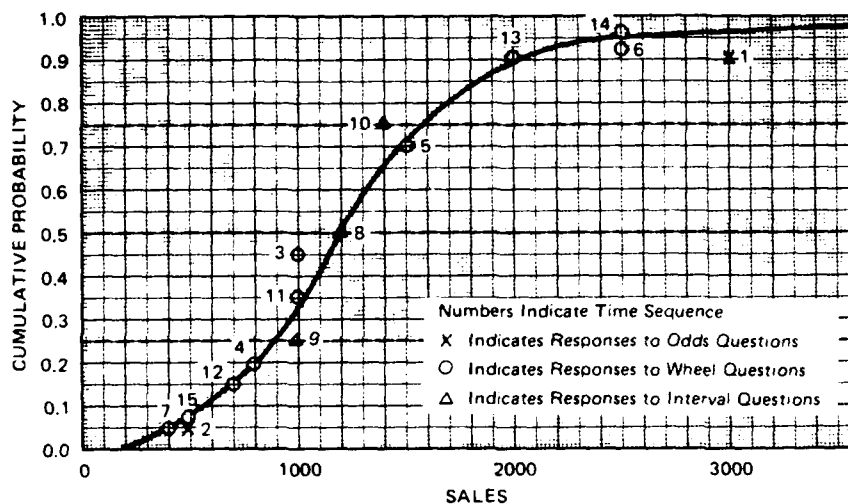
FIGURE 5   EXAMPLE OF A CURVE FITTED TO RESPONSES

5.5 Verifying

In the last stage of the interview, the judgments are tested to see if the subject really believes them. If the subject is not comfortable with the final distribution, it may be necessary to repeat some of the earlier steps of the interview process.

Graphically representing the responses as points on a cumulative distribution and interpreting this distribution (perhaps in terms of a density function) provide an important test and feedback. The interviewer will naturally have to explain how the responses were plotted and how the fitted curve should be interpreted; this generally does not present problems. An examination of the distribution itself cannot show whether or not the distribution agrees with the subject's

judgment. However, it can show implications of the subject's responses and thereby provide feedback. If some responses are not consistent with the subject's judgment, they will have to be modified. A few examples will illustrate this form of verification.

For the first example, assume that the subject has been asked about the number of tank units in use in a specific country 5 years from now. The responses to various questions have been interpreted as points on a cumulative distribution, and these points have been fitted reasonably well by the curve shown in Figure 6(a). The curve satisfies the necessary condition that the probability level increases (or more stringently does not decrease) when the number of units increases; hence, there is no violation of the laws of probability (cf. Section 1.1). However, the shape of the curve may be of some concern in that it first rises, then levels off, and finally rises again. One interpretation of the cumulative distribution is that the probability associated with any interval is equal to the difference in values of the cumulative distribution at the end points of the interval. For instance, the probability of a number of units between 600 and 1,000 is equal to the probability of a number of units not exceeding 1,000, less the probability of the number not exceeding 600 units--i.e., 0.53 - 0.36 = 0.17. At the same time, the interval from 400 to 600 has a probability of 0.27, and the interval from 1,000 to 1,200 has a probability of 0.31. Even though each of the two extreme intervals has only half the width of the central interval, they are both more likely to contain the revealed value of the number of units than is the central interval. In other words, the revealed value is more likely to be found around 500 or 1,100 units than around 800 units, which is a value between the other two. The subject may find this conclusion from the assigned distribution counter to his intuition, and he may therefore want to go back and revise some of his old responses or start all over again. However, the conclusion could also agree with the subject's judgment; he may consider the number of units as depending on whether the tank will work only in the desert, or in both the desert and the tropics. If that is the case, he expects the number of units to be
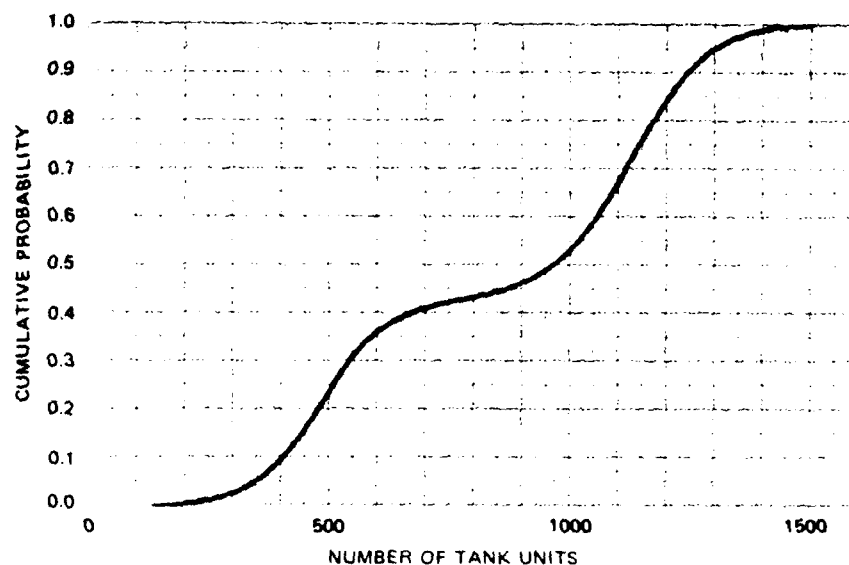
(a) CUMULATIVE DISTRIBUTION
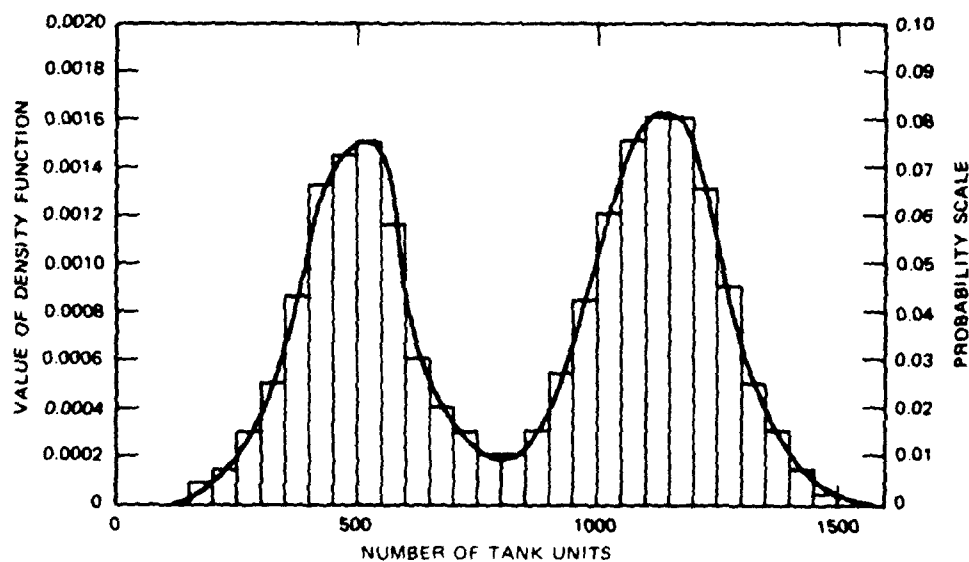
(b) APPROXIMATE DENSITY FUNCTION

FIGURE 6   EXAMPLE OF DERIVATION OF A DENSITY FUNCTION FROM A
CUMULATIVE DISTRIBUTION

around 500 units for the desert only; otherwise, he expects it to be around 1,150 units. The important thing here is to be able to interpret a cumulative distribution and to check it for reasonableness.

It is easier to see the above argument by sketching the density function. This has been done in Figure 6(b). The range of outcomes for the number of tank units has been divided into intervals of 50 units in length. A bar is drawn above each interval with the height of the bar equal to the probability of the interval as read off the cumulative distribution. The total area under all bars is 50 (the width of all intervals). If all numbers on the vertical axis are reduced by a factor of 5C, then the total area is one, and the bar graph is an approximation of the density function. It is easy to fit a smooth curve through the bar graph to represent the density function. It is clear from Figure 6(b) that the distribution is bimodal--that is, it has two peaks.

Another example is shown in Figure 7(a), cumulative distribution, and 7(b), density function. The uncertain quantity is supposed to be the average time between failures for a new component. The cumulative distribution seems to be increasing at a faster pace as the failure time increases. The implication for the density function is that it increases as the time increases up to 110 hours but that there is no probability of the average between failure time exceeding 110 hours. There is a high probability (17%) of the failure time falling between 105 and 110 hours at the same time as it is judged impossible that the failure time exceeds 110 hours. This looks somewhat abrupt, and it is reasonable to discuss this implication with the subject. One explanation may be that because the specifications state 110 hours between failures, the research aims at driving the failure time below this level; at the same time, there is no incentive to improve the reliability beyond specifications. Even in this case, the subject would usually want to assign some probability of exceeding the 110-hour specification.

A second part of the verification process is based on a sequence of pairs of bets. Each pair is chosen so that the two bets would be
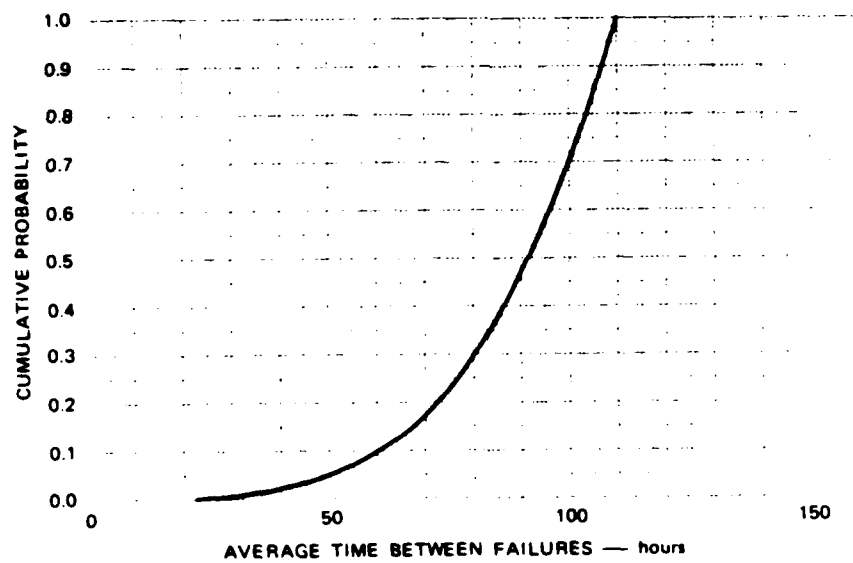
(a) CUMULATIVE DISTRIBUTION
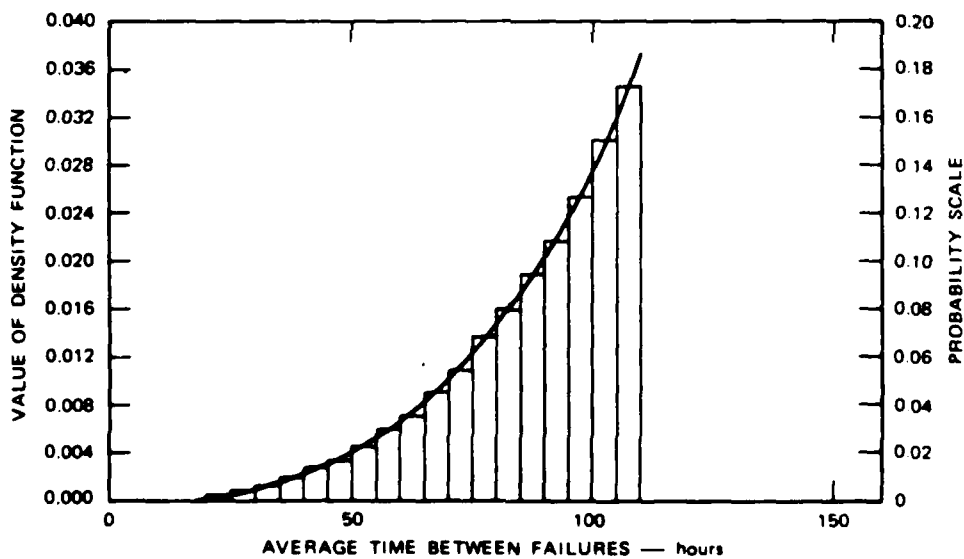
(b) APPROXIMATE DENSITY FUNCTION

FIGURE 7 A SECOND EXAMPLE OF THE DERIVATION OF A DENSITY FUNCTION FROM A CUMULATIVE DISTRIBUTION

52

equally attractive if the curve from the preceding phase is consistent with the subject's judgment. Let us use the curve in Figure 5 for an example. This curve represents the cumulative distribution for sales of a product called THETA. The purpose is to take a point on the curve, which thus shows the probability that sales will not exceed a given value, and construct two bets that have equal value given that probability assignment. For instance, we read from the curve that the probability of the sales not exceeding 1,400 units is 0.65. Using the probability wheel, make the orange sector 65% of the total and then ask the subject whether he prefers to bet on orange at one spin of the wheel or on the sales being below 1,400 units. If he finds it difficult to choose--i.e., if he is indifferent between the two bets--then this confirms the point on the curve.

There should be a few such indifference responses before the process is ended. This provides the subject and the interviewer with confidence that the curve represents the subject's judgment. The final test is to ask the subject if he would be willing to base his own bets in accordance with the plotted curve.

5.6 Length of the Interview Process

A typical interview, in our experience, lasts anywhere from 30 to 90 minutes. The length of the interview depends on many factors, such as the importance and complexity of the uncertain quantity and the subject's previous experience in probability encoding. The pre-encoding steps are more time-consuming than the actual encoding step in about half of all cases. This is particularly true when it is important to understand the structure underlying the subject's judgment and when dealing with subjects who are deeply involved in the project under analysis, especially if they have not had any exposure to the inter- viewer or to the decision analysis effort. The pre-encoding steps alone can in special situations take up to a couple of hours. The encoding step may take up to 1 hour if the quantity is very important,

53

or it may last only 5 or 10 minutes if only a few points are needed on the distribution.

5.7 Other Techniques

The purpose of this section is to comment briefly on other encoding techniques that can be found in the literature but which generally are weak when it comes to practical applications.

It should be clear that the encoding techniques discussed in this manual stress the interaction between interviewer and subject. We find that having the subject assign a probability distribution without the help of an analyst often leads to poor assignments. This is true even for subjects who are well trained in probability or statistics. The main reason for our emphasis on interaction is that it is difficult to avoid serious biases without having an interviewer present.

The technique of having a subject fill out a questionnaire without an interviewer present suffers from the lack of the interaction between *interviewer and subject and usually leads to serious biases.* Questionnaires can be used as a first approximation to the encoding process, but only with subjects that are experienced in probability encoding.

An interactive computer interview can make use of iterative checking techniques, such as the interval technique, and thereby avoid some of the pitfalls inherent in direct response modes. However, the balancing effect of personal interaction is still missing, and the result is almost always centrally biased. Again, we do not recommend using a computer program unless the subject has been through a number of actual interviews that dealt with similar uncertain quantities. Moreover, even when a subject has had a long experience with the computer interview, it should not be used for encoding new types of quantities. However, in situations in which the decision problem is not important enough to justify the cost of having an interviewer perform the interview or when an organization uses probabilities regularly and extensively to communicate about uncertainty, interactive computer

interviews and questionnaires might prove valuable. An example of an interactive probability encoding program is described in Appendix D.

There are procedures that ask the subject for the parameter of a named distribution; e.g., a normal distributuion or a beta distribution. Our experience indicates that subjects will give such parameters, but they usually do not understand the full implications. We consider the choice of named distributions a modeling consideration and believe that it should not be made part of the encoding process.

A simple procedure is to encode only three fractiles such as the 10%, 50%, and 90% fractiles. They may then be fit to a named distribution or be used directly in a decision tree. For example, a normal distribution can be approximated by three steps, using the 10%, 50%, and 90% fractiles with probabilities 25%, 50%, and 25%. The three values are likely to look consistent, but may nevertheless be poor representations of the subject's judgment. For example, simply asking for three fractiles will not reveal whether a central bias might be operating. Even when three values are sufficient as inputs to a model, it would be wise to use a longer encoding procedure to make it more likely that the encoded distribution corresponds to the subject's judgment. It is then a simple matter to obtain an approximation of the encoded distribution that is suitable to the decision model.

6 EXAMPLE OF A SESSION

The following interview is a close simulation of an actual interview. The uncertain quantity and numbers have been disguised. Every real interview differs widely because of the subject, the interviewer, and the quantity for which the uncertainty is encoded. The purpose of this example is to give the reader an impression of what might occur in a real situation rather than to demonstrate all biases or methods of elicitation (all of which would never be found in a single real situation).

The background to the encoding session is the following. Charles Steel has been involved in a decision analysis for the ACME Corporation whether or not they should launch a new product, the THETA machine. He has gone through the deterministic phase and performed sensitivity analyses that showed the yearly sales of THETA at maturity (interpreted as the growth year) as one of the most important state variables. He has learned that Ed Smallcastle from the Marketing Department has worked with similar products in the past and has been conducting some application studies for THETA and could therefore be considered one of the most knowledgeable persons within ACME with respect to evaluating the THETA sales potential. A meeting has been set up with Steel, the interviewer (I), and Smallcastle, the subject (S).

[The purpose of the first part of the session is to develop rapport with the subject. It is important first to get to know him, to make sure he understands the purpose of the session, to note his major worries and concerns regarding the use of his answers, and to assure him that his results will not be misused.]*

---

*The comments within brackets either refer to what the interviewer is doing or to his thoughts about what is happening.

I: Hello Ed. As you know, the purpose of our getting together is to try to get a feeling for your judgment regarding the market potential of THETA. We have talked with a number of people who have been suggested as knowing a lot about this, and one of those whose opinion is valued is you. We are describing the market development by the sales when the product has reached maturity, which should be after around 3 years. The level of sales is just one of the many inputs to the decision analysis which is quite uncertain, and some of our preliminary sensitivity analyses indicate that it is one of the key variables. I am sure this isn't surprising to you. However, as you know there is a lot of money at stake in this product, and before going ahead we are going to try to determine the best estimates from a number of subjects.

What piece of information are you going to base most of your judgment on? What do you see as some of the critical factors involved here?

S: Well, I am thinking about the past successes we have had with our market research. I know it's useful, but I really can't base all my opinion on the market research. Now I <u>really</u> want to know the results of the New York City field tests before I give my opinion on the market for THETA. In fact, I wouldn't want to commit resources until I hear from that.

I: Before we get into that, do you have any questions about what we will do with the results of this discussion?

S: Well, I sort of understand what you guys do with your decision analyses. You kind of fit these numbers in, and you're going to crank them through and get a profit lottery. We'll see how it works. Be kind of interesting.

I: Before going ahead, I want to bring up one thing, though, that has come out in a couple of previous discussions. One of the worries that has been expressed by people is that they will give us their honest judgment and then someone will turn around and make them commit to producing a level where they have only a 30% chance of success. You are in sales, where similar things might happen. Does that bother you?

S: No, not really. I'm pretty used to making commitments like that. In the last four years when I was a salesman, I was in the 400% club every year: I beat my estimate by 400%. I came out on the good side every time.

I: So typically you would tend to underestimate so you have a good chance of being successful?

S: Sure, wouldn't you? That's the way the game is played, isn't it?

I: Yes, that's the way it's played all right. Partially what we are trying to accomplish with the analysis, though, is to put together our best judgment to decide whether or not to play the game at all. And in that case, it can end up that the project doesn't really seem worthwhile if you underestimate, so there is a risk on either side. If you overestimate somebody will turn around and say: hey, how about those high numbers.

S: Well, that's all very fine and good. But you know 3 years ago, when they first thought about this project, they were estimating about 20,000. That's a fantastic number for this thing. And then they chopped it down to 5,000 and now it's down to around 1,000. That kind of thing has got to stop--no more of these optimistic estimates. Let's put down something we know we can get and get the project going.

I: OK, that's what we really want--we'd like you to put down what you really believe, including some of the wild things if you think they have any chance, or the conservative ones if you think there is a chance that it might be as bad as that.

[At this point the interviewer might go into some further details regarding estimating procedures within the company if he believes further time is required before the subject is at ease with the general procedure.]

The next step is to define and structure the Quantity and to elicit the assumptions that the subject is making in thinking about the Quantity.

I: Well, let's move on and talk more specifically about THETA. We are interested to find out what you think of the sales potential of THETA. Let's define that as the level of sales in a year when THETA has reached maturity. By the way, do you agree that that should be after around 4 years?

S: Well, that should be about right for a product of this kind. It normally takes 3 to 5 years, but I'll go along with 4.

I: So, it's clear to you what we mean when we talk about the level of sales at maturity?

S: I take it that we are only talking about domestic sales. It's really uncertain whether we will try to put THETA on the export market.

I: I agree. We can add the export market at a later stage if we want to. Now, how do you go about making your assessment of the sales? Do you break down your estimate in some way?

S: We have made detailed studies of different application areas, of course, but that's very detailed, you know.

I: It's up to you. If you can think of the aggregate market we'll continue to talk about it. But if you'd be more comfortable if we broke the total sales down into sales in the various application areas, then we should do that. We should choose whatever is easiest for you.

S: Taking the total market is fine with me.

I: Well, Ed, tell me, what's a really bad kind of situation that you think might possibly happen?

S: Well, I really think this project ought to go ahead and there aren't too many things that can go wrong. Now I'm behind this project.

I: Let me ask it in another way. How many orders do you think that you have in the bag or you are sure of getting?

S: Oh, I've got 100 orders in my drawer right here, and I'm sure if you really went out and beat the bushes, you know, there would be no sweat at all--almost no matter what price you charge--300 or 400. No problem at all!

I: You think your sales at maturity are going to be 300 or 400?

S: Sure, I think we will make that.

I: What about a wild guess on the high side?

S: Just a minute! You are assuming that we are going ahead with this project, and you know we may not. It might be

all very fine to go ahead, but we may not get management approval for this thing until I don't know when. You know--what kind of assumptions are you putting behind this? How much money are we going to get? How did R&D turn out? Does the thing work? Did we get the automatic timing device working? You know, you are asking me an awful lot of questions here.

I: Well, let's not get down into a lot of detailed assumptions here. I'll want to set down specific assumptions in a little while--but first I just want to get a general impression. Certainly you can assume we go ahead, and let's assume we go ahead in the beginning of next year. So you have a product, it works, it's been field tested, it's been shown, it's reliable, just as you would expect--and let's say you go out at the beginning of next year. Now in that situation, what do you think is a really high number?

S: Oh, there are some guys around here, you know; they would talk 5,000 or 10,000--I would actually cut that down to around 2,000. That's at the outside.

I: Now Ed, assume that you're downstream 4 years from now and you have probably reached maturity and at that point someone tells you that you got 2,000. OK? That is a really high estimate. What would be the main reasons that something like that could have happened?

S: Oh well, you know something that high--there are a few reasons like suppose the THETA market continues its growth, like some people think it might. And if you really got management support behind this, and other magnetic devices don't come through like some people hope they will. I suppose it's possible, particularly if we get that field test and it is successful. I guess--2,000--it could be done but I wouldn't want to count on it in any way. We certainly couldn't allocate any of the company's resources to wild estimates like that.

I: What kind of surprises do you think could happen? Of course, you wouldn't want to make decisions on them, but what kind of things might happen to push it above that.

S: Well, you know, some guys talk about--oh, they've got some wild ideas about using THETA for things like inventory control. You know, it just might. Hell, you know, I can't even visualize the market for it. We can't justify the program on the basis of those markets; we have to work with proven markets. It really depends on the size of the

61

accounting market and how many THETA's we can sell--that sort of thing. Going off on somebody's latent markets--it's too chancy.

I: So you say the main reason for going above 2,000 would be outside of the accounting market. Are you assuming in your estimates that the sales are all in the accounting area?

S: Yes, of course, that is what our program is for.

I: Well, the estimates we are trying to get are really completely inclusive, and it might be worthwhile for us to try and do them separately. Do you want to first estimate the accounting market and then talk about what kind of outside latent markets there might be?

S: Oh, I don't know. Whatever you want to do.

I: Let's try to include the latent markets.

[By now the interviewer has determined two key motivational biases that are likely to overshadow most other biases: the subject wants management to continue the project, and he wants to give a low estimate because underestimation would tend to increase his perceived credibility. The interviewer needs to help in setting those biases aside. The interviewer is also suspecting a cognitive central bias due to availability of previous estimates and market research. The interview is now carried into the conditioning phase.]

I: When you are thinking about an uncertain quantity, it often helps to think about it as a member of a broader class of similar quantities. If you are estimating the reliability of a new machine, you may relate it to the reliability of similar machines that have been developed within the past. If you are estimating the cost of a development program, you may think of the deviation of cost from budget for other development programs. We are now discussing the sales of THETA. Can you think of such a reference class in connection with THETA?

S: Well, it can be related to the sales of some similar machines that we have developed in recent years. But they have, of course, been quite different in size. How do I compare them?

I: You can perhaps relate actual sales to what the original forecast was, I mean at a stage similar to where you are

62

today with THETA. That way you don't have to think explicitly about the sizes of the various markets. How do actual sales compare with the original forecast, on the average? And how much can sales vary relative to the forecast for products of this kind?

S: That's a tough one. You know, some of these forecasts are unreal. R&D believes they've got a fantastic new product, and it turns out to be a flop on the market. But, on the average..., well, maybe the actual market turns out to be above the forecast about half the time and below half the time.

I: And how much can it vary around the forecast?

S: If it really takes off, and it sure did with OMEGA, it could be four times as high. And a flop might not give you more than a tenth of what you had predicted.

I: Now let's return to THETA. How good a product do you think it will be?

S: It's a fairly advanced product. It should be doing quite well.

I: What does "quite well" mean in terms of sales at maturity?

S: I'd say around 1,000.

I: How did you get to that number?

S: Well, that's a nice, round number, isn't it? And it's been kicked around in our plans for a while now.

I: Are there any other numbers that have been used in connection with THETA?

S: I guess that depends on who you are, but there have been forecasts between 500 and 2,000.

I: But your own estimate is close to the official forecast.

S: It seems all right.

[The subject's estimate does not seem to be very extreme and there is no need to evaluate the predictability of the outcome of the sales. There may be a tendency to fall back on an "established" forecast. The use of extreme scenarios increases the availability of other

possibilities to the subject and should reduce the anchoring effect of that forecast.]

I: Well, we discussed a high value for the sales volume a little while ago. Let's return to that for a moment. If you include the latent markets, what's a really high number for you? One, that you would be surprised if the sales turned out to be greater than that number.

S: Oh, I think if we can include those markets I would say 3,000 or so.

I: Well, what kind of odds would you give me that the actual placements five years from now are going to be above 3,000?

S: #'#'!!!!! I'd never commit to a number like that.

I: This is between us. I don't think you ought to. Let's just say you have full support; if I gave you 10 to 1 odds on it--do you think you would take it?

S: Anything my boss is behind, I'm behind.

I: What are you saying? You could have one chance in ten of making that?

S: Well--yes, sure. You know if it were my own business--but we have an organization here, we have had some tough situations in the last year, and a lot of products came out, and they didn't meet up to specs. We have to have one or two that come out where we beat our market estimates, and then we get our reputation back. In the present situation, we hate to commit to a number like that.

I: Look, I'm not talking commitments at all now. What I'm trying to figure out is what the odds are; obviously you are not going to commit to something where you don't have at least a 50-50 chance of making it, but let's not even worry about that because commitments shouldn't really come out of a probability distribution.

S: Well, you know between us girls I would give it 1 chance in 10.

I: OK, let's turn to the low side. What would be a similarly low number? Don't forget to include the latent markets; we are talking about total THETA sales at maturity.

S: Well, I presume you already said we are going ahead--presume the assumption is we are going to start off early next year. Well, probably the worst thing that can happen is that some of these other devices they're talking about get the jump on us, and people decide really not to make the investment in THETA. Then some of these orders we've got in the drawer--they just might change their mind.

I: There is an existing market already, isn't there?

S: Yes, sure. It's been growing nicely.

I: So if you include all the applications, what might be a low sales volume?

S: I'd say around 500.

I: Suppose I told you for a fact that less than 500 units were sold. Could you give me a scenario that is consistent with this?

S: Well, if you include those other applications--and they are pretty well independent of the accounting market--I really can't see much of anything replacing that. Include those and it still could probably do 500, even if the accounting market for this thing goes sour.

I: What I meant with a scenario, Ed, is one specific set of events that actually could happen and end up in less than 500 orders. And just give me one example.

S: Oh, well, a cut back in our marketing staff, and they only give me about 20 salesmen. You know these guys are running around doing other things too. And if at the same time the THETA market doesn't grow, and also we don't have the quality on the machine that we are really hoping for--we have some timing device problems--and our price--you haven't mentioned much about price, but let's say we had to price it high. That would keep the orders down there. Well, let's see, 500 units; that works out to be, oh, about 40 a month. I think 20 guys can sell 40 a month. That sounds reasonable. If we had a bigger staff we could do more, but they would never cut me down to below 20 staff and still make the project go.

I: Well, let's assume that you are going to get full marketing support on this thing as long as you can demonstrate that there is more market available because it looks like a profitable unit. And now let's make some very specific assumptions on this. Later on we can talk

about how it would differ if it weren't on those assumptions. First of all the machine is produced and is on spec. OK. As for the specs you guys wrote--they're met all the way. Let's talk strictly about demand at maturity.

Now--we want to get down to some numbers. I would like you to include in your judgment such factors as possible ways the THETA market might go and different ways the competition might act.

S: What do you mean about competition? I presume you mean Flextronics? OK, I'll think about it.

I: The IOTA product also comes through on schedule and you get full support from the sales force as I mentioned before.

S: When will the product be available? You were talking about next year before--we've really got to get cracking on developing the market right now with test sales calls and so on. I'd have to get started 3 or 4 months from now, and I'm not sure that's coming.

I: Let's just assume it is--OK? Later on we'll talk about what if it were delayed; but first let's assume introduction next year, and you get the signal to go ahead early enough to really prepare yourself. What are the odds that you end up with sales below 500?

S: Well, they are fairly small. I'd give 1 in 20.

[The interview is now moving into the quantification of judgment. The interviewer has chosen to start with odds on the extremes, he will then use the wheel and check with interval questions. The interviewer should continue to be sensitive to motivational and cognitive biases. The interviewer now introduces the probability wheel and sets it at 20% orange and 80% blue.]

I: Ed, if I gave you a chance to play an interesting game here--

S: We are really getting down to business now--I heard about your wheel.

I: Here is the game, Ed. We are going to spin this wheel and if it ends up pointing to the orange you win $1,000; if it ends up pointing to the blue, you get nothing. Do you want to play?

S: Sure do. Can't lose either way.

I: Well you can either bet on the wheel--on the orange--or you can bet on THETA sales ending up above 1,000 units.

S: I'll take THETA sales--I can easily beat 1,000--I don't like the wheel.

[The interviewer moves the wheel up to 80% orange.]

I: How about it now, Ed?

S: Now I'll take the wheel, of course.

I: Now somewhere in between there, you changed your mind. Let's try this.

[The wheel is now set at 35% orange.]

S: Which side am I betting on, the orange or the blue?

I: Here is the game again: If you bet on the wheel, you win if it's on the orange, if you bet on THETA, you win if sales are above 1,000.

S: Boy, that one's a lot tougher. I don't know. I'll have to think about that.

I: Are you willing to decide on the bet by flipping a coin at this point?

S: No, I think I'll still bet on THETA.

[Now the wheel gets moved to 45% orange.]

I: How about now, Ed?

S: That's pretty close. I guess at this point I would flip a coin.

I: OK, now let's try another one. This time you will win on THETA if the sales are below 800.

S: Oh, I don't like to bet on that kind of thing.

I: Do you think there is a good chance that it could happen?

S: Well, I wouldn't mind a target that low, but I don't think we could get the program approved.

I: Oh. Can you think of any situation where you had a target that looked pretty easy to get but where something happened so that you didn't make it even then.

S: Yes, our last two or three programs went that way.

I: So now let's try for you to be in the following role. You make the estimates and whatever game we play, just assume that you get promoted out of your job into a different part of the company, so you can't really change it anymore, and all you have to think about is just watching. What do you think? Now you watch a game where you either bet on below 800 or bet on the orange. Which would you rather do?

[The wheel is still set at 45% orange.]

S: Oh, I'd take the wheel.

[The wheel is changed to 25% orange.]

I: How about now, Ed?

S: I'd still take the wheel.

I: Ed, the wheel is now set to about 20%.

S: I'd flip a coin about here.

[While this interview is going on, points corresponding to responses are being plotted on a graph paper as shown in Figure 8. The horizontal axis represents sales, and the vertical represents probability. At this point 4 points have been plotted, corresponding to a 5% probability of placements below 500 from the second odds question, 20% probability below 800 from the last response using the wheel; 45% below 1,000 from the previous response using the wheel, and 90% below 3,000 from the first odds question. This is being done outside the view of the subject. By looking at that plot, it is obvious that some more points are needed in the range of 1,000 to 3,000.]

I: OK. Ed, let's focus in on 1,500 units. At this time, do you bet on below 1,500 or on the wheel.

[The wheel is set to about 60% probability.]

S: You mean I win if sales are below 1,500?

68

I: Yes.

S: I'd bet on THETA below 1,500.

INTERVIEWER __CHARLES STEEL__   SUBJECT __ED SMALLCASTLE__   DATE __6/14/78__

Numbers Indicate Time Sequence

X Indicates Responses to Odds Questions

O Indicates Responses to Wheel Questions

VARIABLE ____THETA SALES AT MATURITY____
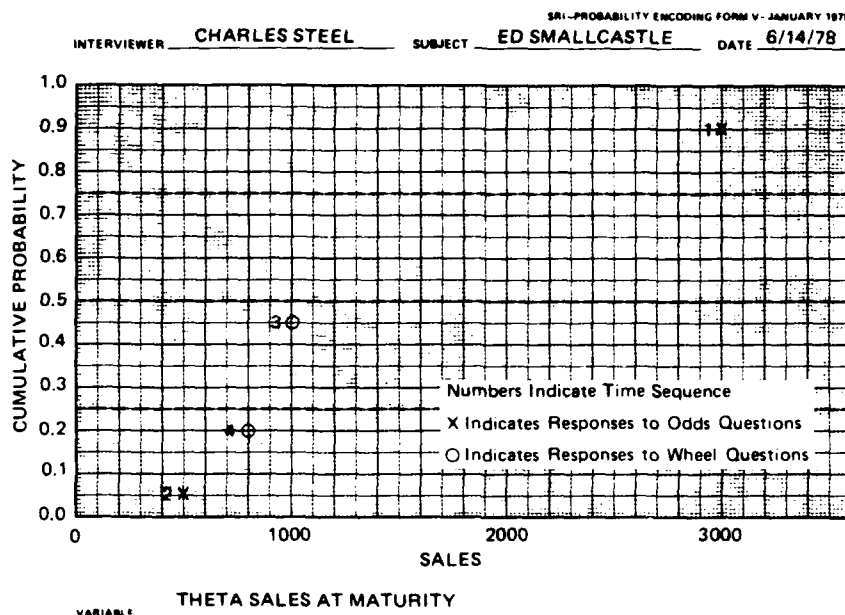
FIGURE 8   INITIAL PROBABILITY RESPONSES

I: OK, I'll move the wheel up a little further--

S: Stop right about there.

[The wheel is now about 70% orange.]

I: Let's try a high one, Ed. How about 2,500 units and this
    time you bet on above 2,500.

[The wheel is set at about 30% orange.]

S: I'll take the wheel. You can make it a lot smaller than
    that.

I: I move the wheel down to 15% orange.

S: Further, and a little further yet. There is fine.

[The wheel was stopped at 8%.]

I: Let's try a real low one. This time you get the bet above
     400 units or on the wheel.

S: That's a good one. You will have to go really high on the
     wheel, I'd say over 90% orange.

I: OK, here's 90%--does that look about right?

S: A little more orange--yeah--about there is fine.

[The wheel stopped at about 95% orange. At this point, the interviewer
has 7 points on the plot (see Figure 9). The first in consistency has
also been detected: The probability of sales less than 3,000 is smaller
than the probability of sales less than 2,500. The interviewer now
proceeds by using interval questions to try to determine the consistency
of this plot.]
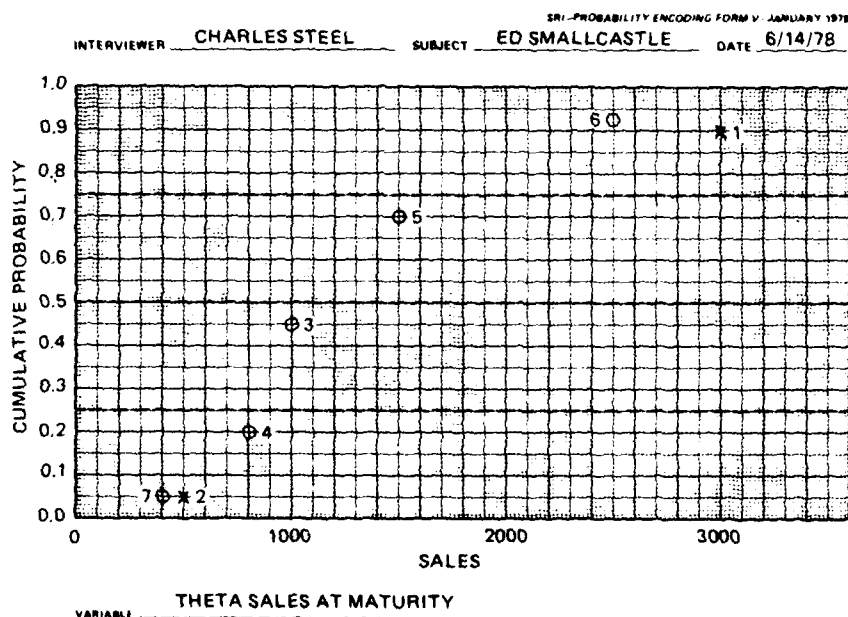
FIGURE 9   RESPONSES TO ODDS AND WHEEL QUESTIONS

I: We're going to put the wheel away now—I know you like it, but let's try another approach. The game is still for a hypothetical $1,000. I'll divide the range of all possible outcomes of the sales into two ranges and ask you which of the two ranges you would prefer to bet on. The first time you can either bet on sales above 1,000 or sales below 1,000. Which would you rather bet on?

S: Above 1,000.

I: How about above or below 1,200?

S: That one is tougher. About there—I could take either one.

I: OK, now let's define the two ranges as sales being less than 700 and sales being between 700 and 1,200. Which one of these two ranges would you prefer to bet on?

S: I'd bet on sales being between 700 and 1,200.

I: How about anywhere between 900 and 1,200, or below 900?

S: I have a feeling it's going to be between 900 and 1,200, so I'll pick that one now.

I: OK—How about between 1,000 and 1,200, or below 1,000?

S: Well you are starting to put me in a box there—somewhere around there I'd switch, I guess.

I: OK, let's change the game to two ranges above 1,200. You can either bet on sales falling between 1,200 and 2,000 or above 2,000. Which range would you rather bet on?

S: What two ranges?

I: 1,200 to 2,000, or 2,000 and above?

S: 1,200 to 2,000.

I: OK, 1,200 to 1,500 or 1,500 and above?

S: Why don't you try 1,200 to 1,400 and then I'll switch.

I: OK, now try these two ranges. Range one is from 1,000 to 1,400, outside that range is the other one. Would you rather bet on the inside or on the outside?

S: They are about the same to me.

[The subject is definitely exhibiting a central tendency (central bias) in the answers to the interval questions. This is clear from the plot (see Figure 10). A curve drawn through points 8-10, representing the responses to the interval questions, would be much steeper than a curve drawn through points 3-7, representing the responses to the wheel questions. This means that the first curve represents a narrower distribution. The interviewer at this point has two alternatives: to confirm his check points with some other technique, or to explain to the subject his bias on this kind of question and try to train him to improve in his responses. In this interview, an attempt will be made to confirm the check points with the wheel.]

I: Based on all the answers you gave me so far, I'll show you on the wheel what your graph says. Let's just see if that agrees with your judgment or if you want to make some further adjustments. Would you rather bet above 1,000 or on the orange?

SRI–PROBABILITY ENCODING FORM V JANUARY 1978
INTERVIEWER CHARLES STEEL   SUBJECT ED SMALLCASTLE   DATE 6/14/78

VARIABLE THETA SALES AT MATURITY
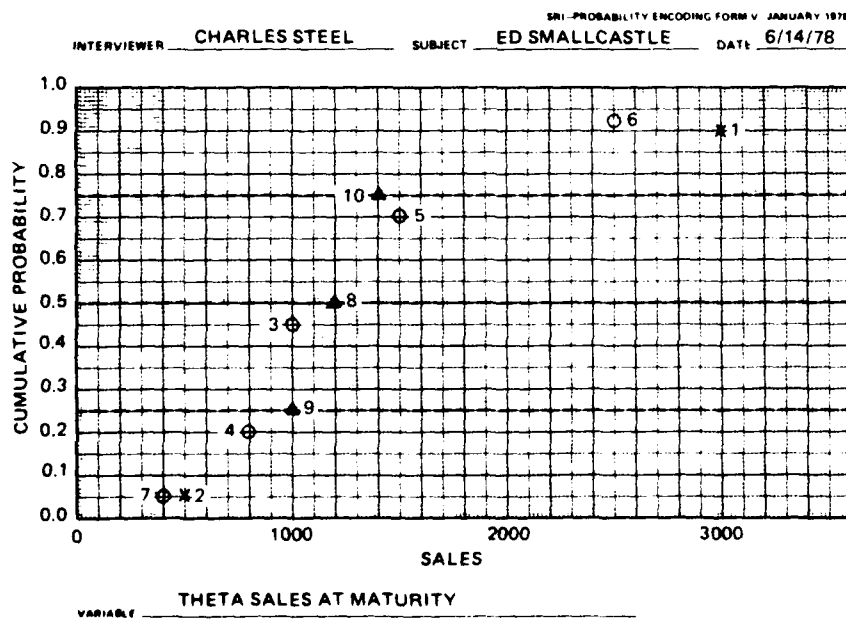
FIGURE 10   RESPONSES TO ODDS, WHEEL, AND INTERVAL QUESTIONS

[The wheel is set at 60% orange.]

    S: That's pretty good, I'll take the wheel.

[The wheel is changed to 65% orange.]

    S: I'm indifferent there.

    I: How about below 700 or betting on the orange?

[The wheel is set at 15% orange.]

    S: I'd say that's pretty close to my indifference point.

    I: OK, one more here--on above 2,000, or on the orange?

    S: I'll bet on the orange.

    I: How about now?

[The wheel is set to approximately 10% orange.]

    S: About there is fine.

    I: Let's try out on the ends. Above 2,500?

    S: Is that about 5%?

    I: Yes.

    S: I'll take it.

    I: You mean you are close to indifferent between betting on the wheel and on the sales?

    S: Yes.

[Because of the decision problem involved, the decision depends much more heavily on the range below 800 (the break-even point for this product is 500 units). Therefore, we would like to confirm one or more points in the lower range.]

    I: Which would you rather bet on, below 500 or on the orange?

[The orange is now set at about 20%.]

    S: You would have to make it much smaller.

I: How about this?

[The wheel is now set at 10% orange.]

S: That's close...maybe just a little smaller.   That's fine.

[The answer was around 7-8%. The interviewer plots the latest point and then draws a smooth curve that fits the points in the plot, the later points in particular.]

I: Well, Ed, I think we have enough points here. Let me show you what we did. This is the plot that I made (see Figure 11).   Each of those points represents one of your responses--like the last thing you said was about an 8% probability below 600.   Look at this one way up here at 3,000.   When I first asked you--way back when--what are the chances that it will be above 3,000, you said one out of ten, so there would be a 10% probability above, or a 90% probability below 3,000.   Then later, based on the wheel, you came to a 95% probability below 2,500.   That's a pretty strong inconsistency.   But our experience tells us that those early odds questions are unreliable; the responses often overstate the outside probability.   I would guess, at this point, that this last point that we got at 2,500 better reflects your belief.
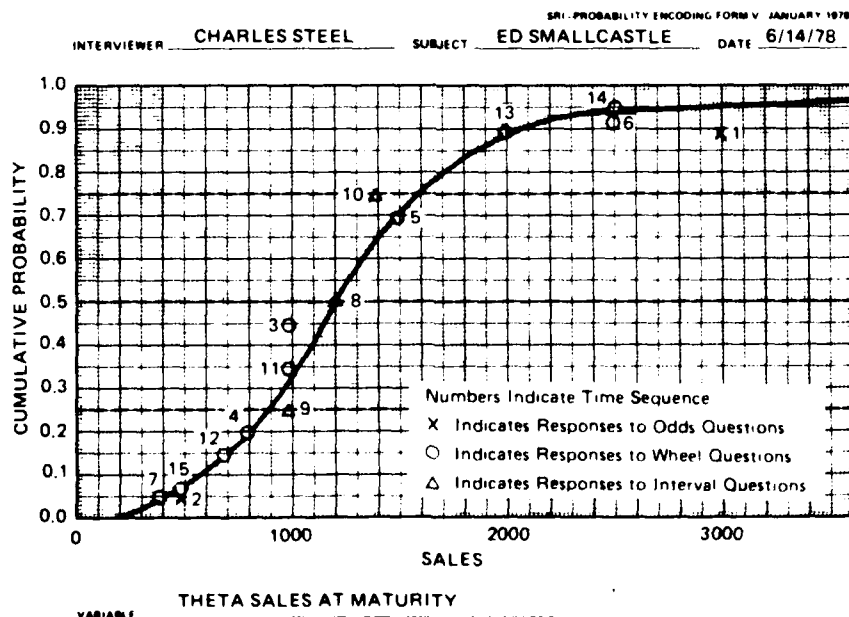
FIGURE 11   FINAL CURVE FITTED TO RESPONSES IN THE ENCODING PHASE

74

Let me just test that. Which would you rather bet on: above 3,000 or on the wheel when it's now set at about 5%?

S: I'd bet on the wheel.

I: Well, what this really means is that the odds above 3,000 rare less than one in twenty rather than one in ten that you said earlier.

S: Yes, I think I'd change my mind. Being above 3,000 isn't that likely.

I: OK, do you have any other questions before I leave here?

S: Well, you have been interviewing a lot of other guys around here. What did they say? I'd like to see their curves. What are you going to do with mine anyhow? You have to be careful how you use them--this kind of thing could be interpreted the wrong way.

I: Well, we're getting all these curves because this particular input to the decision analysis that we are doing is really critical. Now it turns out that the decision problem doesn't depend that much on the high side--it's more in the range of 400 to 1,000 that we are really sensitive to the estimate, and what we'll be doing is getting you and four other guys together in a meeting to talk about the differences in your estimates. Before that, though, we want to get everybody individually, so they think it through all by themselves. Do you have any reservations about doing that?

S: No, none at all--sounds like a lot of fun. But before we do that, will you do me one favor? Could you get my boss' curve? He's the one who's really got to commit himself on this project.

I: Funny you mentioned that, Ed, because he's the guy who says he's going to base his estimate just about completely on what you say.

S: That's good to hear!

[The interview resulted in a distribution that is obviously dependent on a whole list of assumptions regarding management support, timing, quality of product, and such. Shifts in the distribution for changes in these assumptions were encoded later.]

7 ADDITIONAL TOPICS

7.1 Discrete Distributions

The encoding techniques discussed so far may not work well (perhaps not at all) when the number of possible outcomes for the uncertain quantity is small. The prime interest rate 3 months from today is an example of such a quantity; it is usually measured in quarter percentage points and only moves a small amount. The purpose in this section is to show how the encoding techniques can be modified for such quantities.

In the case of a discrete set of outcomes, the probability distribution can be described by a **mass function** that shows the probability associated with each outcome. [See Figure 12(a) for an illustration.] It can also be given in the form of a **cumulative distribution** [Figure 12(b)] as in the continuous case. The cumulative distribution increases only at values that represent possible outcomes. A **histogram** is an alternative form of a mass function. Each probability is then represented by a bar located at the corresponding outcome [Figure 12(c)].

The probability wheel can be used as a P-method as before. It can also be used to assign probabilities to individual outcomes and thereby provide consistency checks.

The interval technique will no longer produce a value for the median, but rather will produce two inequalities (e.g., the probability of the prime rate being less than 5-1/2% is less than 50%, and the probability of the prime rate being less than or equal to 5-1/2% is greater than 50%. The best use of the interval technique is in the verification phase.

The subject can be asked to assign relative likelihoods (or odds) to two events. For example, the subject may state that it is twice as likely that the prime rate will be 5-1/4% than that it will be 5%. This

77
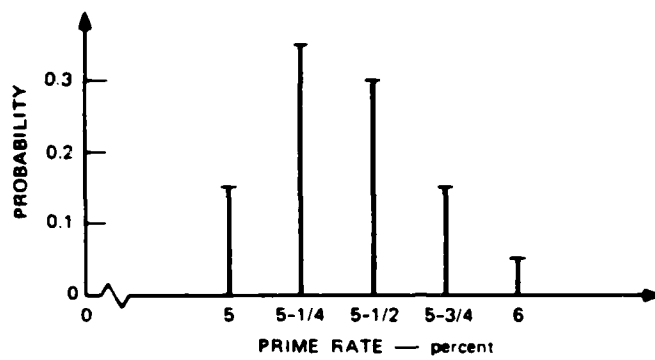
(a) MASS FUNCTION
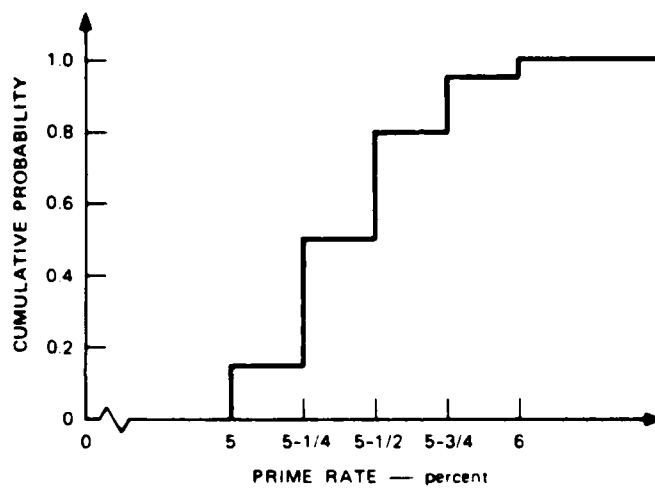
(b) CUMULATIVE DISTRIBUTION
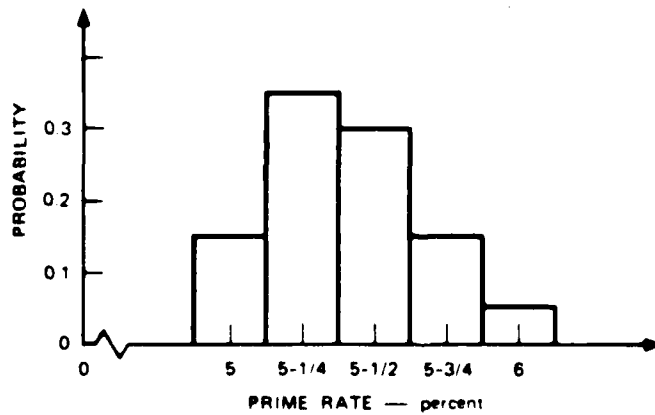
(c) HISTOGRAM

FIGURE 12   THREE REPRESENTATIONS OF A DISCRETE
PROBABILITY DISTRIBUTION

will produce a point on the cumulative distribution if either of the two events must occur. Otherwise, it will provide a consistency check.

To summarize, the probability wheel remains the most useful technique for the quantification of judgment, supplemented by a few odds questions or direct assignments. The interval technique and odds questions should be used mainly in the verification phase.

7.2 Rare Events

In many applications, one unlikely outcome of a crucial state variable has a significant effect on the total result. The reliability of a production process may be an example where malfunctioning could be very expensive. A nuclear power plant is an extreme example with respect to the high cost and the low probability of an accident occurring. The quantification of judgment is difficult for rare events; it is also difficult to present or display small probabilities.

Rare events present special problems in probability encoding since the standard techniques do not work well for small probabilities. Subjects often find that it is difficult to discriminate between sizes of small sectors on a probability wheel. Similarly, the interval technique is more effective with the central part of the distribution even though theoretically it can also be used to generate the tails of a distribution. For example, continuing to split the lowest interval into equal parts generates an event with a probability of roughly 0.001 after ten steps. However, the final response is the composite of ten different responses and even slight biases in the responses lead to substantial error when compounded ten times.

As mentioned previously, fixed probability events such as poker hands or coin-tossing sequences can be used as external reference events for low-probability events. One can also develop reference processes that can serve as P-methods, at least when it comes to discriminating between orders of magnitude for the size of a probability. An example of such a technique is to show the subject a chart divided into squares 1,000 x 1,000--that is, 1 million squares in all. It is easy to make

such a chart from standard graph paper. The subject is asked to imagine that each of the 1 million squares has equal probability of being selected by some random mechanism. The event that a particular square will be chosen then has a probability of 10^{-6}, which is small enough for almost any rare event that might practically be encoded. Reference events with other probabilities are defined by selecting the relevant number of squares.

However, our experience with probability encoding for rare events indicates that probabilistic modeling is generally more effective than direct encoding. For example, for an event to occur, it may be necessary that a sequence of other events occur. These intermediate events may not be low-probability events and standard encoding procedures can then be used. The problem of encoding the probability of a single rare event is thus transformed into the task of modeling the probabilistic structure of a sequence of events and then encoding the larger probabilities of these events.

7.3 Accuracy, Honesty, and Calibration

Three factors affect the 'goodness' of a probability assignment. One factor is the subject's knowledge of the problem area. We use the term accuracy to represent the closeness between a probability distribution for an uncertain quantity and the actual outcome. The other two factors are the motivational and the cognitive biases. We use honesty as a concept representing lack of motivational bias. The probability assignment that is void of any motivational bias is said to be honest. Similarly, calibration is used to represent the degree of cognitive bias. Probability assignments that agree completely with the subject's judgment are said to be perfectly calibrated.

It would be useful to be able to separate the effects of accuracy, honesty, and calibration in that knowledge about such effects would help to train subjects in probability encoding. For instance, a subject who produces honest and well-calibrated probability assignments needs no

further training. It may be possible to separate accuracy and calibration because one refers to expertise in the problem area and the other refers to lack of cognitive biases, but it is impossible to isolate the degree of honesty. A revealed bias can be motivational or cognitive (or both), and there is no way to tell unless further information is provided.

If we assume, however, that a subject's probability assignments are honest, we can display his calibration through a _calibration function_. Consider, for example, a collection of a subject's probability assignments of 0.20 to many different events. We can check in each case whether the event actually occurred and can then calculate the relative frequency of such occurrences. With perfectly calibrated assignments, we can expect a relative frequency around 0.20, at least in the long run. However, experience shows that a subject is more likely to show a different relative frequency. We can now plot the relative frequency for each probability level, and the curve thus obtained is the calibration function. Perfectly calibrated assignments would lead to a calibration function with the relative frequency everywhere equal to the probability level. _Figure 13 shows a typical calibration function, with the perfect calibration function at the 45° line. It may be noted that some observations usually fall outside the entire range; thus, the value at probability zero is greater than zero, and the value at probability one is less than one._

Motivational biases may be eliminated or reduced by the use of a reward structure that encourages honesty. Such structures, which are sometimes called _scoring rules_, assign a score to a probability assignment in the light of the revealed value of the uncertain quantity. It is doubtful, however, whether they will have much effect because a motivational bias usually arises from an implicit and much stronger reward structure within the organization. A scoring rule can also be used to evaluate probability assignments, but cannot separate the effects of accuracy, honesty, and calibration.
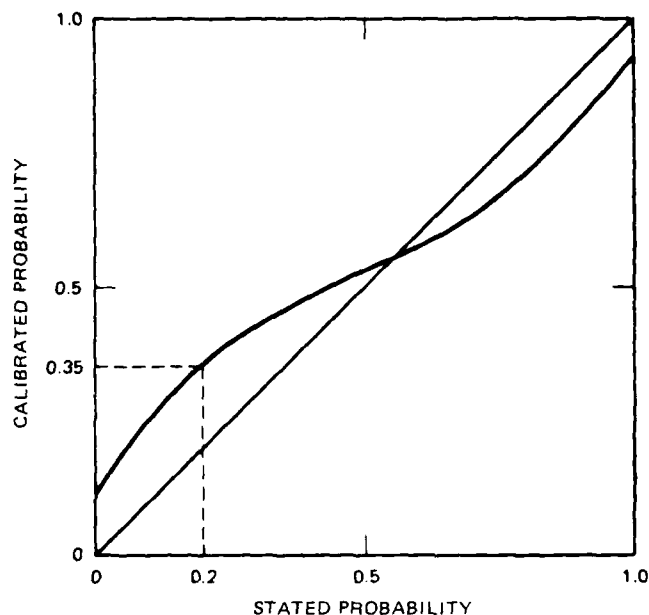
FIGURE 13   A TYPICAL CALIBRATION FUNCTION

7.4   Training Versus Calibration

If we know a subject's calibration function, we can calibrate any probability assignments he makes. For instance, if the curve in Figure 13 represents a subject's past performance, then we might want to use a probability of 0.35 as our probability assignment if he has just assigned a probability of 0.20.   It should be stressed that the knowledge about past performance should be used like any other piece of information; one thing this means is that the decision maker will use it only to the degree he finds it relevant. For example, the subject may have gone through an extensive training program, which invalidates data before the training.

However, there are some definitional and technical problems in using the calibration function to correct a probability distribution.

For example, does information about probability assignments of 0.20 also provide information about assignments of 0.80, the complementary probability? We prefer to use the calibration function as an indicator that a problem of cognitive bias exists. The solution to that problem is training.

Training should consist essentially of feedback about past performance. Most subjects make few probability assignments during their work (meteorologists may be exceptions); therefore, they do not receive extensive feedback. It is then possible to provide training through experimental sessions wherein the subjects are asked about any kind of uncertain quantities. Because we want to improve their ability to quantify judgment, it is immaterial whether we use almanac questions (such as "What was the legal whiskey production in the U.S. in 1970?") or quantities relating to their field of expertise.

We feel that it is important in the long run to have subjects who will understand probability encoding procedures and who by training make well-calibrated assignments. In practice, we almost invariably use training rather than calibration because the subject would object to the interviewer, or even his superior, changing his judgment.

7.5 Use of Multiple Subjects

It is not uncommon for the decision maker to have access to more than one subject for judgment about some uncertain quantity. The subjects can be expected to assign different distributions because they have different states of information. The decision maker may be satisfied with a collection of probability distributions for use as a basis for forming his own judgment. He may also p' 'er to have the subjects reach a consensus that he may use either directly as input to the decision analysis or as a basis for his own judgement. The question is then how to reach a consensus that in some sense is best.

A consensus can be reached only if there is an exchange of information among the subjects. The exchange can take one of two forms: either the persons meet as a group to exchange information, or they

1.0

2.8   2.5

3.2   2.2

3.6

4.0   2.0

1.1

1.8

1.25   1.4   1.6
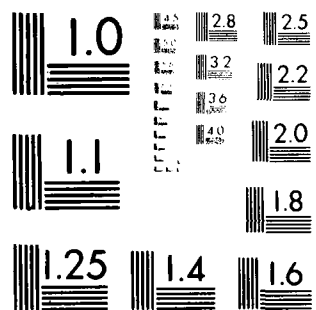
MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS 1963 A

communicate anonymously. The latter form is supposed to reduce the influence from group members with dominant personalities. It is similar to the Delphi method that originally was related to single-number estimates. The procedure is simply to feed to every other member of the group the distributions assigned by each person. This new information provides the option to revise their probability distributions. Repeating this formal feedback procedure a few times usually leads to some convergence. The method can be improved by permitting the people to include arguments as to why their distributions may differ from those of the others.

We have found it more productive to let the subjects eventually get together to trade information rather than maintaining anonymity. This makes their states of information, and thereby their probability assignments, more similar--if they are well calibrated. An interviewer may serve as a moderator to reduce the influence of certain strong personalities if necessary. The sources of disagreement can be detected more easily with a group discussion than with a formal feedback procedure. For example, the group members may find that they agree on a model structure but disagree on a particular input to the structure. Work toward a consensus is easier after they have found out exactly where they disagree.

We have tested both procedures in seminars and experiments; the subjects have generally found that they have gained more from discussions. We thus recommend group discussion rather than formal feedback procedures, but some caution must be taken when using either one. Most important is that the subjects are given an opportunity to think through the problem and that their individual distributions are encoded before the exchange of information begins. The exchange may be done in different ways: it may be unstructured with everyone discussing until nobody has anything left to say, or it may be conducted by an interviewer who tries to bring out the essential arguments from the subjects. The choice depends on the individual situation.

84

A BIBLIOGRAPHY OF ENCODING PROBABILITY DISTRIBUTIONS

1. Introduction

The purpose of this bibliography is to present a guide to the extensive literature related to encoding probability distributions. (The first version of this bibliography was written in 1971; this revision includes several items published in the period 1971-1978, but does not claim to be exhaustive.) It is divided into two parts: the first concerns individual encoding and the second relates to resolution of multiple expert opinion. Included are theoretical as well as experimental works. The selection has been affected by the considerations discussed below.

The emphasis of the bibliography is on the encoding of distributions for uncertain quantities with more than two possible outcomes; nevertheless, some works related to encoding of individual probabilities have been included. References on resolution of multiple expert opinion have been restricted to works discussing how an internal consensus might be reached by the group as opposed to works presenting nonbehavioristic aggregation methods.

Even though the literature is extensive, few works discuss encoding procedures for practical applications. The paper by Spetzler and Staël von Holstein (1975) is an exception, but it summarizes only the main points of an earlier version of this manual. The book by Brown, Kahr, and Peterson (1974) includes an extensive section on probability encoding. Compared with this manual, it is more concerned with the actual encoding phase than with the pre-encoding interaction between interviewer and subject. The same also applies to the chapter on probability encoding by Peterson et al. (1972).

Two articles may serve as the best references to the relevant psychological literature. Hogarth (1975) provides an extensive review of the overall literature. The article by Tversky and Kahneman (1974) summarizes their own research on modes of judgment and biases, and has influenced the methodology presented in this manual.

2. Individual Encoding

Numerous studies exist on the encoding of individual probabilities (which is equivalent to encoding probabilities for dichotomous quantities). Only a few references will be included here, and they will generally be of the nature of an overview because this bibliography is primarily concerned with probability encoding for nondichotomous quantities. However, direct encoding of points on a cumulative distribution function would be equivalent to repeated encoding of individual probabilities; hence, such studies might be relevant here also.

A scoring rule is an incentive scheme for eliciting honest probability assignments and could thus be regarded as an encoding technique. However, only a few works will be included that give the essence of their usefulness in probability encoding. Their references will help locate the remainder of the works in the area.

2.1 Mainly Theoretical

Brown, R. V., A. S. Kahr, and C. R. Peterson, Decision Analysis for the Manager (Holt, Rinehart, and Winston, New York, 1974).

Includes an extensive section on probability encoding that compares various encoding techniques and discusses what constitutes a good probability assignment.

de Finetti, B., "Does it Make Sense to Speak of 'Good Probability Appraisers?'" in The Scientist Speculates--An Anthology of Partly-Baked Ideas, I. J. Good, ed., pp. 357-364 (Heinemann, London 1962).

Characterizes "good probability assignments" and presents the quadratic scoring rule as a means of encouraging honest assignments.

_____, "Methods for Discriminating Levels of Partial Knowledge Concerning a Test Item," <u>The British Journal of Mathematical and Statistical Psychology</u>, Vol. 18, pp. 87-123 (1965).

Discusses the use of probabilistic responses for eliciting partial knowledge about test items together with ways of scoring them.

Fox, B. L., "A Bayesian Approach to Reliability Assessment," Memorandum RM-5084-NASA, The Rand Corporation, Santa Monica, California (1966).

Suggests the use of the mode and a dispersion measure for assignments related to the parameter of a Bernoulli process.

Hampton, J. M., P. G. Moore, and H. Thomas, "Subjective Probability and Its Measurement," <u>Journal of the Royal Statistical Society</u>, Ser. A., Vol. 136, pp. 21-42 (1973).

Literature review on encoding techniques.

Hogarth, R. M., "Cognitive Processes and the Assessment of Subjective Probability Distributions," <u>Journal of the American Statistical Association</u>, Vol. 70, pp. 271-289 (1975).

Extensive review of recent research on judgmental processes for the encoding of subjective probability distributions.

Huber, G. P., "Methods for Quantifying Subjective Probabilities and Multi-Attribute Utilities," <u>Decision Sciences</u>, Vol. 5, pp. 430-458 (1974).

Reviews encoding techniques that have been empirically studied.

Kahneman, D., and A. Tversky, "Intuitive Prediction: Biases and Corrective Procedures," in a Special Issue of <u>Management Science</u> on Forecasting Methodologies and Applications, Wheelwright and Makridakis, eds. (in press).

Presents an approach to elicitation and correction of intuitive forecasts. It is closely related to Appendix A of this manual.

Lin, C. Y., and G. J. Schick, "On-Line (Console-Aided) Assessment of Prior Distributions for Reliability Problems," <u>Annals of Reliability and Maintainability</u>, Vol. 9, pp. 13-19 (1970).

The procedure uses successive subdivisions and/or direct assignments of points on a distribution function.

Matheson, J. E., and R. L. Winkler, "Scoring Rules for Continuous Probability Distributions," <u>Management Science</u>, Vol. 22, pp. 1087-1096 (1975).

Discusses the use of scoring rules for continuous probability distributions.

Morrison, D. G., "Critique of: 'Ranking Procedures and Subjective Probability Distributions,'" Management Science, Vol. 14, pp. B253-B254 (1967).

Criticizes Smith (1967) and advocates the use of "successive subdivisions."

Murphy, A. H., and R. L. Winkler, "Scoring Rules in Probability Assessment and Evaluation," Acta Psychologica, Vol. 34, pp. 273-286 (1970).

Discusses suitable properties for scoring rules to be used in probability assignment.

Peterson, C. R., C. W. Kelly III, S. Barclay, and T. H. Hazard, "Probability Distribution for a Continuum," a preliminary chapter for Handbook for Decision Analysis, Engineering Psychology Programs, Office of Naval Research (1972).

An extensive example of a probability encoding session that demonstrates different encoding techniques.

Pratt, J. W., H. Raiffa, and R. Schlaifer, Introduction to Statistical Decision Theory (preliminary edition) (McGraw-Hill, New York, 1965).

Advocates the use of "successive subdivisions" and presents some approximate procedures for cases when the distribution can be assumed to belong to a certain parametrizable family.

Raiffa, H., Decision Analysis (Addison-Wesley, Reading, Massachusetts, 1968).

Presents the technique of using "successive subdivisions" in the form of a dialogue between a decision analyst and his client (Section 7.3).

_____, "Assessment of Probabilities," Harvard University, unpublished manuscript (1969).

Discusses external validity (calibration) of subjective probability assignments as well as strictly proper scoring rules together with criteria for choosing among such rules.

Savage, L. J., "The Elicitation of Personal Probabilities and Expectations," Journal of the American Statistical Association, Vol. 66, pp. 783-801 (1971).

Derives the general class of strictly proper scoring rules by considering probabilities as special cases of rates of substitutions; also presents applications.

Schlaifer, R., _Analysis of Decisions Under Uncertainty_ (McGraw-Hill, New York, 1969).

Includes a chapter on the encoding problem.

Selvidge, J., "A Three-Step Procedure for Assigning Probabilities to Rare Events," in _Utility, Probability, and Human Decision Making_, D. Wendt and C. Vlek, eds., pp. 199-216 (D. Reidel, Dordrecht, Holland, 1975).

Presents a procedure for the assignment of probabilities to rare events.

Smith, L. H., "Ranking Procedures and Subjective Probability Distribuitons," _Management Science_, Vol. 14, pp. B236-B249 (1967).

Presents a mathematical (but unrealistic) encoding technique. See critiques by P. E. Green and D. G. Morrison in the same issue.

Spetzler, C. S., and C.-A. S. Stael von Holstein, "Probability Encoding in Decision Analysis," _Management Science_, Vol. 22, pp. 340-358 (1975).

A summary of an earlier version of this manual.

Stael von Holstein, C.-A. S., _Assessment and Evaluation of Subjective Probability Distributions_, Economic Research Institute, Stockholm, Sweden (1970a).

Reviews literature on encoding techniques. Includes the experiments in Stael von Holstein (1971a, b, c; 1972a).

_____, "Measurement of Subjective Probability," _Acta Psychologica_, Vol. 34, pp. 146-159 (1970b).

A review paper that is essentially a condensed version of Stael von Holstein (1970a).

_____, "A Tutorial in Decision Analysis," Stanford Research Institute, unpublished manuscript (1972b).

Includes an introduction to probability encoding in the context of decision analysis.

_____, "The Continuous Ranked Probability Score in Practice," in _Decision Making and Change in Human Affairs_, H. Jungermann and G. de Zeeuw, eds., pp. 263-273 (D. Reidel, Dordrecht, Holland, 1977).

Presents an approximation that makes a scoring rule for continuous probability distributions usable in practice.

Winkler, R. L., "The Assessment of Prior Distributions in Bayesian Analysis," _Journal of the American Statistical Association_, Vol. 62, pp. 776-800 (1967a).

A presentation of a number of encoding techniques together with an experimental study. Leads to a questionnaire for use in probability encoding.

_____, "The Quantification of Judgment: Some Methodological Suggestions," _Journal of the American Statistical Association_, Vol. 62, pp. 1105-1120 (1967b).

A thorough presentation of the use of scoring rules and other payoff schemes to elicit honest assignments.

2.2 _Mainly Experimental_

Alpert, M., and H. Raiffa, "A Progress Report on the Training of Probability Assessors," Harvard University, unpublished manuscript (1969).

An experiment with almanac kinds of questions for which subjects assigned five fractiles. Shows that assignments generally are too narrow.

Brown, T. W., "An Experiment in Probabilistic Forecasting," Draft Report R-944-ARPA, The Rand Corporation, Santa Monica, California (1972).

An experiment on 6-month forecasts for which subjects assigned seven fractiles.

Edwards, W., "The Theory of Decision Making," _Psychological Bulletin_, Vol. 51, pp. 380-417 (1954).

A review of research on behavioral decision theory. Includes experiments with assignments of individual probabilities.

_____, "Behavioral Decision Theory," _Annual Review of Psychology_, Vol. 12, pp. 473-498 (1961).

A continuation of Edwards (1954).

Kahneman, D., and A. Tversky, "Subjective Probability: A Judgment of Representativeness," _Cognitive Psychology_, Vol. 3, pp. 430-454 (1972).

Representativeness is a mode of judgment; accordingly, the probability of an event or a sample is evaluated by the degree to which it is representative of the major characteristics of the process or population from which it originated.

_____, "On the Psychology of Prediction," _Psychological Bulletin_, Vol. 80, pp. 237-251 (1973).

Presents more experimental evidence of representativeness as a mode of judgment.

Leonardz, B., and C.-A. S. Stael von Holstein, "A Comparison Between Bayesian and Classical Methods for Estimating Unknown Probabilities," Project ORBS Technical Report No. 3, Division of Applied Mathematics, Brown University (1967).

An experiment with oddly shaped dice generating Bernoulli events. Subjects assigned the mean and a 95% credible interval.

Lichtenstein, S., B. Fischhoff, and L. D. Phillips, "Calibration of Probabilities: The State of the Art," in _Decision Making and Change in Human Affairs_, H. Jungermann and G. de Zeeuw, eds., pp. 275-324 (D. Reidel, Dordrecht, Holland, 1977).

Reviews the experimental literature on calibration.

Luce, R. D., and P. Suppes, "Preference, Utility, and Subjective Probability," in _Handbook of Mathematical Psychology_, Vol. 3, R. D. Luce, R. R. Bush, and E. Galanter, eds., pp. 249-410 (John Wiley and Sons, New York, 1965).

Includes descriptions of experiments with encoding of individual probabilities.

Peterson, C. R., and L. D. Phillips, "Revision of Continuous Subjective Probability Distributions," _IEEE Transactions on Human Factors in Electronics_, HFE-7, pp. 19-22 (1966).

A probability revision experiment concerning the parameter of a Bernoulli process. Subjects assigned the .33 and .67 fractiles.

Quinn, D. J., and J. E. Matheson, "The Use of Judgmental Probability in Decision Making," SRI International, Menlo Park, California (1978).

Includes a section on the validation of probability assessments with references to experiences from practical applications.

Rapoport, A., and T. S. Wallsten, "Individual Decision Behavior," _Annual Review of Psychology_, Vol. 23, pp. 131-176 (1972).

A review of research on decision behavior covering the period from 1965 through 1970.

Schaefer, R. E., and K. Borcherding, "The Assessment of Subjective Probability Distributions: A Training Experiment," _Acta Psychologica_, Vol. 37, pp. 117-129 (1973).

Assignments concerning proportions of students with given characteristics. The encoding procedure used fractile assignments as well as hypothetical samples.

Selvidge, J., "Assigning Probabilities to Rare Events," Harvard University, Graduate School of Business Administration, unpublished Ph.D. dissertation (1972).

Reports results of interviews with decision makers in two fields concerned with rare events: specialty insurance underwriting and nuclear safety analysis. Supplemented by experimental results and a suggestion for an encoding procedure.

Slovic, P., "From Shakespeare to Simon: Speculations--and Some Evidence--About Man's Ability to Process Information," _Oregon Research Institute_, Research Monograph, Vol. 12, No. 12 (1972).

A review of some internal procedures that people use when making judgments and decisions. It incorporates some of the results given by Kahneman and Tversky (1972, 1973) and Tversky and Kahneman (1973).

Slovic, P., B. Fischhoff, and S. Lichtenstein, "Behavioral Decision Theory," _Annual Review of Psychology_, Vol. 28, pp. 1-39 (1977).

A review of research on decision behavior covering the period from 1971 through 1975.

Stael von Holstein, C.-A. S., "The Effect of Learning on the Assessment of Subjective Probability Distributions," _Organizational Behavior and Human Performance_, Vol. 6, pp. 304-315 (1971a).

An experiment with oddly shaped dice generating Bernoulli events. Subjects assigned median and quartiles.

_____, "Two Techniques for Assessment of Subjective Probability Distributions - An Experimental Study," _Acta Psychologica_, Vol. 35, pp. 478-494 (1971b).

A sequel to Stael von Holstein (1971a). The encoding procedure also used four hypothetical samples.

_____, "An Experiment in Probabilistic Weather Forecasting," _Journal of Applied Meteorology_, Vol. 10, pp. 635-645 (1971c).

Assignments concerned with temperature and precipitation. The distributions were formulated as sets of three to eight probabilities.

_____, Probabilistic Forecasting: An Experiment Related to the Stock Market," _Organizational Behavior and Human Performance_, Vol. 8, pp. 139-158 (1972a).

Assignments concerned changed in buying prices over 14-day periods for shares quoted on the Stockholm Stock Exchange. The distributions were assigned in the form of five probabilities.

Tversky, A., and D. Kahneman, "Availability: A Heuristic for Judging Frequency and Probability," _Cognitive Psychology_, Vol. 5, pp. 207-232 (1973).

Explores a mode of judgment by which a person evaluates a probability by the ease with which relevant information is recalled or imagined.

_____, "Judgment Under Uncertainty: Heuristics and Biases," _Science_, Vol. 185, pp. 1124-1131 (1974).

A comprehensive review of the work on modes of judgment and biases. Some aspects are described in more detail in Kahneman and Tversky (1972, 1973) and Tversky and Kahneman (1973).

_____, "Causal Schemes in Judgments Under Uncertainty," in _Progress in Social Psychology_, M. Fishbein, ed., Lawrence Erlbaum Associates, Hillsdale, in press.

Advocates that the impact of evidence on intuitive judgments of probabilities depends critically on whether it is perceived as causal, diagnostic, or incidental.

Winkler, R. L., "The Quantification of Judgment: Some Experimental Results," _Proceedings of the American Statistical Association_, pp. 386-395 (1967c).

Assignments concerned with the point spread in American football games. The spread was divided into six intervals to which probabilities were assigned.

_____, "Probabilistic Prediction: Some Experimental Results," _Journal of the American Statistical Association_, Vol. 66, pp. 675-685 (1971).

An extended analysis of Winkler (1967c).

See also de Finetti (1962), who mentions an experiment with the outcomes win, lose, or draw in Italian soccer games, and Winkler (1967a), who

93

relates an experiment with assignment of probability distributions for unknown proportions.

3. Resolution of Multiple Expert Opinion

Many references are concerned with formal aggregation of distributions assigned by more than one person, and they are given by Winkler (1968). The only work that has dealt with behavioristic group assignment procedures (i.e., procedures by which the persons receive feedback from each other and then revise their assignments) is Winkler (1968). However, this bibliography includes some references to the Delphi method, although this has so far only dealt with point estimates, and references to general group decision studies, that also may be relevant to the reconciliation of expert opinion. Morris (1971, 1974) is also of some interest here in that he discusses the resolution problem as an instance of Bayesian inference.

Brown, B., "Delphi Process: A Methodology for the Elicitation of Opinions of Experts," Report P-3925, The Rand Corporation, Santa Monica, California (1968).

A description of the Delphi method.

Brown, B., S. W. Cochran, and N. C. Dalkey, "The Delphi Method, II: Structure of Experiments," Memorandum RM-5957-PR, The Rand Corporation, Santa Monica, California (1969).

Dalkey, N. C., "The Delphi Method: An Experimental Study of Group Opinion," Memorandum RM-5888-PR, The Rand Corporation, Santa Monica, California (1969).

Presents the results of a large experiment with the Delphi method. The design of the experiment is presented in the preceding reference.

Hogarth, R. M., "Methods for Aggregating Opinions," in Decision Making and Change in Human Affairs, H. Jungermann and G. de Zeeuw, eds, pp. 231-255 (D. Reidel, Dordrecht, Holland, 1977).

A review of models and methods for aggregating opinions.

Kelley, H. H., and J. W. Thibaut, "Group Problem Solving," in The Handbook of Social Psychology, Vol. 4, G. Lindzey and E. Aronson, eds., second edition, pp. 1-101 (Addison-Wesley, Reading, Massachusetts, 1969).

Maier, N.R.F., "Assets and Liabilities in Group Problem Solving: The Need for an Integrative Function," *Psychological Review*, Vol. 74, pp. 239-249 (1967).

Morris, P. A., "Bayesian Expert Resolution," Stanford University, Department of Engineering-Economic Systems, unpublished Ph.D. dissertation (1971).

Presents a Bayesian solution to the problem of how the decision maker should revise his prior distribution after having received assignments by one or more experts. It does not discuss the assignments that would have to be made in practical applications.

_____, "Decision Analysis Experts Use," *Management Science*, Vol. 20, pp. 1233-1241 (1974).

Introduces the basic methodology of Morris (1971).

Wallach, M. A., N. Kogan, and D. J. Bem, "Group Influence on Individual Risk Taking," *Journal of Abnormal and Social Psychology*, Vol. 65, pp. 75-86 (1962).

Winkler, R. L., "The Consensus of Subjective Probability Distributions," *Management Science*, Vol. 15, pp. B61-B75 (1968).

A thorough discussion of methods for combining subjective probability distributions--some mathematical and some entailing feedback and/or group discussion. Supplemented by experimental results.

Winkler, R. L., and L. L. Cummings, "On the Choice of a Consensus Distribution in Bayesian Analysis," *Organizational Behavior and Human Performance*, Vol. 7, pp. 63-76 (1972).

An experiment wherein subjects played the role of decision makers faced with a number of experts' probability distributions for some uncertain quantity.

Appendix A

INTUITIVE JUDGMENT: BIASES AND CORRECTIVE PROCEDURES

Daniel Kahneman, University of British Columbia
Vancouver, British Columbia, Canada
and
Amos Tversky, Stanford University
Stanford, California, U.S.A.

A.1 Introduction

Decision analysis is a formal framework for analyzing complex decisions involving uncertainty. It consists of a coherent set of logical and statistical procedures that are applied to the data which characterize the decision under study. These data consist of hard facts such as resources and prices as well as subjective judgments that express the beliefs and the values of the decision maker. Although we attempt to substitute objective facts for subjective judgments whenever possible, most decision analyses contain a significant judgmental component.

In the absence of objective procedures for measuring the probabilities of unique events and the utilities of nonmonetary outcomes, we treat the subjective judgments of the decision maker (or an expert who is acting on his behalf) as measurements of uncertainty and value. These judgments often contain essential information, but they are usually fallible and often biased.

Biases of judgment are classified into two types: motivational and cognitive. Motivational biases refer to (conscious or unconscious) distortions of beliefs and values motivated by one's personal interests or prior commitments. For example, an expert who is opposed to the development of nuclear power plants may overestimate the likelihood of a nuclear accident, whereas an expert who has a vested interest in this industry is likely to underestimate the probability of such an accident.

97

Often unwittingly, people attempt to influence decisions by slanting estimates upward or downward. Motivational biases are common when people provide estimates of their own future performance. Under some circumstances, such estimates are likely to be overoptimistic--e.g., when several firms are competing for a contract, or when an individual is attempting to project an image of confidence and competence. On the other hand, an individual is likely to provide an underestimate of his future performance whenever his estimate can be viewed as a commitment to achieve a particular result. When estimates are not clearly distinguished from commitments, a salesman will tend to underestimate future sales, and a production manager will tend to overestimate the time required to complete a particular job.

In addition to the motivational factors, there are several cognitive factors that operate in a more subtle and usually unconscious manner to produce systematic errors or biases. The cognitive biases observed in the intuitive assessment of probabilities and values are analogous to the perceptual biases and illusions observed in the intuitive estimation of distance, for example. In both cases, people do not have a precise method for computing probability or distance. Instead, they rely on certain intuitive methods that usually lead to reasonable estimates. To survive together, drivers and pedestrians have learned to estimate distances and speed with considerable precision. Likewise, the security analyst has learned to estimate the likelihood of success of various business enterprises with reasonable accuracy.

Although the intuitive methods used to estimate distance and probability are generally useful, they often lead to severe and systematic errors. For example, the apparent distance of an object is determined in part by its clarity. The more sharply the object is seen, the closer it appears to be. The rule is quite useful because in any given scene the more distant objects are seen less sharply than nearer objects. However, reliance on this rule leads to systematic errors in the estimation of distance. Specifically, people overestimate distances when visibility is poor because the contours of objects are blurred. On

the other hand, people underestimate distances when visibility is good because objects are seen more sharply.

Three features of this example are worth noting. First, people are not generally aware of the rules that govern their impressions; they are normally ignorant of the important role of clarity in their perception of distance. Second, people cannot deliberately control their perceptual impressions. A foggy mountain looks far away even if one has learned of the effect of fog on the perception of distance. Third, it is possible to learn to recognize the situations in which impressions are likely to be biased and deliberately to make appropriate corrections. In making a decision to climb a mountain, for example, one should consider the possibility that the summit is further than it looks if the day is particularly clear.

A similar analysis applies to the assessment of probabilities and values. As in the perceptual example, people are usually not aware of the basis of their impressions, and they have little deliberate control over the processes by which these impressions are formed. However, they can learn to identify the rules that determine their impressions and to make appropriate allowance for the biases to which they are susceptible.

In the following sections, we describe some common sources of bias and error in the assessment of probabilities and values, and propose a set of procedures designed to elicit the best information that is available to the subject. For easy reference, recommended procedures are labeled by an asterisk (*). The following terms are adopted: the interviewer is the person who is in charge of eliciting estimates and preferences; the subject is the informant who supplies the answers. Our recommendations assume that the information is obtained in a face-to-face interview, but they are readily adapted to the construction of questionnaires to be answered by an expert working on his own.

A.2 The Control of Motivational Biases

The factors that produce motivational biases in estimates are generally well understood by most people, and we shall not elaborate on them. What should perhaps be emphasized is that motivational biases of judgment are not simply lies. In many cases, the biased subject really believes in his erroneous opinions, and the interviewer's task is to help him reach a more balanced conclusion, rather than to "find him out." In other situations, the subject may be under pressure from his organization to slant his answers in a particular direction, and it becomes the interviewer's responsibility to structure the interview and the record so as to protect the subject from pressure, while eliciting unbiased estimates from him. The following procedures are useful to reduce the effects of motivational biases.

(*)Whenever possible, the analysis should rely on impartial subjects who have nothing to gain by slanting their estimates and who have no motive to influence the decision one way or another.

(*)When uncommitted subjects cannot be found, as is often the case, it is the interviewer's responsibility to obtain estimates from subjects who hold divergent views.

(*)Whenever there is a suspicion of motivational bias, the interview should be particularly probing and detailed. The interviewer should stress that he is interested in a comprehensive view of all the factors that affect the problem, and that the subject's reasoning and his ability to assess all relevant factors are as important as his numerical estimates of a particular quantity. It is especially useful to ask for specific arguments and for the details of the subject's reasoning; these are written down and become part of the record. Writing down the subject's comments is a powerful source of motivation for the subject to cover all aspects of the problem and not to neglect obvious arguments against his favored view because it reminds the subject that the quality of his reasoning is part of the record and is subject to criticism and review. (Written notes are probably more effective than tape-recording in this context.)

100

(*)Structuring the problem in terms of fine-grained estimates of several quantities is likely to be more useful than requiring the subject to provide a single global estimate. The requirement to explain each estimate leads the subject to reveal the details of his thinking about the problem. This approach is more effective than cross-questioning the subject about a single global estimate. The interviewer should allocate much more time and more detailed preparation to each interview when there is a suspicion of motivational bias than when such a bias is unlikely.

(*)When an individual is to provide an estimate of his own future performance, much of the discussion should be focused on contingencies that are not under the individual's control but may affect his achievements. The emphasis on external conditions tends to remove the confusion between estimates and commitments or promises. Commitments and promises are often implicitly contingent on certain assumptions about the circumstance under which a job will be carried out—assumptions that normally remain implicit. Explicit discussion of the assumptions and of what is likely to happen if they are not met tends to relieve the individual from the pressures that bias his estimates.

A.3 Remarks on the Elicitation of Estimates and Probabilities

For the purpose of decision analysis, the decision maker, or a subject who acts on his behalf, is asked to express his state of information regarding some events or quantitties that may affect the outcome of the decision. The state of information with respect to a discrete event that may or may not occur (e.g., war between Greece and Turkey before 1980) is summarized by assigning a probability to that event. The state of information with respect to an uncertain quantity (e.g., the revenue of a particular firm next year) is expressed by a probability distribution over that quantity.

Probability distributions provide a more precise and flexible manner of conveying both the subject's knowledge and his uncertainty

about the problem than the commonly used language of estimates. A statement such as, "I think the product will be developed in about 18 months," is misleading and incomplete. Misleading because it implies a degree of certainty that may not be justified and that the subject may not in fact experience; incomplete because it provides no hint about the degree of risk that may be involved in acting on that estimate. The explicit assessment of risks and uncertainties is one of the foundations of the decision-analytic approach.

It is a psychological fact, however, that most people who have not had enough training in decision analysis or in the philosophy of modern statistics find the language of estimation more natural and congenial than the language of probability. It is easier for most people to state the value they 'think' will occur than to assign a probability to a value they do not think is the correct one. The device of using a betting language is quite helpful because people can be induced to bet on events they consider unlikely if the odds are right. The use of this device is explained elsewhere in the manual, and we will not elaborate on it here. The introduction of a betting language, however, is not sufficient to guarantee that the odds chosen by the subject accurately express his knowledge and ignorance about the uncertain quantity that is to be assessed. To extract the best information from a subject, it may be necessary to coax him gently from the estimation language that most naturally reflects his thinking to the language of probability that is most appropriate to the purposes of decision analysis.

A similar compromise may be required with respect to the interpretation of the language of probability. The modern philosophical approach interprets probability as an expression of a subjective state of belief, which is assumed to be reflected in one's betting odds. Most people, however, naturally think of the probability of an event as a property of the outside world rather than as a subjective state of belief or as a betting preference. The elicitation techniques that are designed to extract the information available to a subject must be adapted to this common interpretation of probability.

102

The main objective in interviewing a subject is to obtain an accurate and unbiased expression of his state of knowledge. The sequence of stages in the elicitation interview is an important determinant of the quality of this assessment because any opinion to which the subject commits himself affects all his subsequent opinions. The following sections describe the psychological considerations that apply to each stage in a particular interviewing sequence. The proposed sequence of stages appears to be applicable to many assessment problems, but we specifically wish to avoid the suggestion that it is the only appropriate one. We break down our discussion according to the following stages:

- Structuring of the assessment problem. In this phase, the subject and the interviewer are expected to jointly develop a precise definition of the problem and an approach to it that will be as convenient as possible for the subject and designed to be precise and unbiased.

- Preliminary discussion of the quantity to be assessed, leading to a rough estimate of its magnitude.

- Assessment of the probability distribution over the quantity.

In the following sections we discuss common errors and biases that may arise in each of these stages, and describe procedures that the interviewer may use to avoid or reduce these errors.

A.4 Structuring the Estimation Problem

The first stage of the assessment interview is devoted to achieving a precise definition of the quantity for which a probability distribution is required and of the approach that is to be taken in assessing that quantity. Some important choices must be made in this stage. Suppose we are concerned with next year's revenue of a given firm. We could assess this quantity directly, or we could decompose it in various ways. For example, we could break down sources of revenue by clientele; we could break down revenue in terms of manpower, utilization factor, and income per unit time; or in terms of past revenue and rate

103

of growth. The different ways of assessing revenue bring different data to bear on the problem and could therefore yield different results.

Clearly, some decompositions of information are likely to be better than others. To illustrate, suppose we are asked to assess the number of hens in the U.S. today. As we have little direct information about this quantity, it is natural to express it in terms of other quantities that are easier to assess. Thus, we may estimate the number of hens in terms of the number of farms in the United States and the average number of eggs laid by a hen per day. Although the former decomposition appears more natural, the latter is likely to produce better results because we have more information about food consumption and egg-laying habits than about the number of farms and the number of hens in a farm.

The following considerations should be kept in mind in the process of structuring an estimation problem.

(*)Whenever possible, the subject should be consulted about the appropriate decomposition of the problem. He should be asked to choose the units with which he is most comfortable (e.g., percent increase or absolute value), and the decomposition that allows him the best opportunity to bring his expertise to bear.

(*)When time allows, and for problems of special importance, alternative decompositions should be used and the outcomes compared.

(*)It is useful to decompose an estimation problem in such a manner that the various subproblems are not affected by the same sources of error. From this point of view, a multiplicative decomposition (e.g., number of eggs consumed X eggs produced per hen per day) is often preferable to aggregative decomposition (e.g., breakdown by type of fowl or by size of farm) because the components of the aggregate are susceptible to similar biases of estimation. People are often uncomfortable with multiplicative decompositions because they recognize that a range of uncertainty of 1:3 for each of two components yields a 1:9 range for the final estimate. However, this analysis merely reveals the extent of the uncertainty that actually exists, while other

104

approaches may allow the subject to retain the illusion that his estimates are accurate when in fact they are not.

(*)It is not usually possible to obtain probability distributions for each component estimate. For most purposes, it will suffice to obtain careful assessments of a "best estimate" and a crude measure of the range of uncertainty for each component. The subject can then be guided to consider these values in producing a single probability distribution for the quantity with which he is concerned.

(*)In structuring the assessment of a future quantity (e.g., next year's revenue) it is important to discuss at an early stage the main factors that could affect this quantity and cause it to take extremely high or low values. The main object of this discussion is to encourage the subject to hold in mind a comprehensive model of the uncertain situation and of the extreme outcomes to which some unexpected combinations of factors may lead. It is not necessary to obtain detailed numerical statements for these factors.

A.5 Biases of Estimation

A.5.1 Nonregressive Prediction

People most often derive intuitive predictions and estimates of uncertain quantities from a general schema or implicit model of the situation. The subject who assesses the revenue of a given firm, for example, has in mind a model of the firm, its competitors, and the relevant markets; hence, he selects as his best estimate the value that appears most representative of this model. This mode of judgment, however, is not generally compatible with the principles of statistical prediction because the most representative value is rarely the best prediction. The following example illustrates this mode of judgment and the bias to which it leads.

Consider the prediction of the sales of a new book. The editor who reviewed the manuscript prior to its publication was favorably impressed. He said, "This novel reads like a best-seller. It is as

good as any we have published in the last 3 years." If the editor is now asked to predict sales, he will probably predict that this book will sell as many copies as the most successful book published by the company in the last three years. This value of predicted sales is representative of the editor's impression of the book.

This mode of prediction is common, but it is unsound because it fails to take uncertainty into account. The editor will surely admit that the prediction of book sales is notoriously inaccurate and that the success of a book could depend on many unforeseen factors. Because the value of sales that he predicts is high, such unforeseen factors are more likely to decrease rather than increase sales. In the presence of uncertainty, predictions should always be more moderate than the impressions on which they are based. A reasonable prediction for the sales of a book should fall somewhere between the most representative value based on one's impression and the average sales for books of this type.

Perhaps the most basic principle of statistical prediction is that the extremity of predictions (i.e., the degree to which they depart from a relevant average value) should be controlled by the degree of predictability (i.e., the achievable predictive validity). If predictability is nil, then the same value (e.g., the mean of the relevant class) should be predicted in all cases. If predictability is moderate, one is entitled to depart from the mean in the direction suggested by one's impression. One's prediction should match one's impression only when predictability is perfect (i.e., when there is no uncertainty regarding the quantity in question). In general, intuitive predictions do not obey this principle. Experts and laymen alike often make extreme predictions on the basis of information whose reliability and predictive validity are known to be low. Such predictions are called "nonregressive."

The fallacy of nonregressive prediction is easily demonstrated when one predicts the result of a repeated performance or a replication. The laws of chance indicate that a very high score on the first trial is

106

likely to be followed by a somewhat lower score on the second trail, whereas a very poor score on the first trial is likely to be followed by a relatively higher score on the second trial. Thus, if we examine a group of firms that did exceptionally well last year we will probably find that, on the average, their current performance is somewhat disappointing. Conversely, if we select firms that did poorly last year, we will find that, on the average, they are doing relatively better this year. This phenomenon, known as regression toward the mean, is a mathematically necessary consequence of the presence of uncertainty. The best prediction for a repeated performance of an individual, a product, or a company must be less extreme (i.e., closer to the average) than the original score. Psychological research has shown that intuitive predictions consistently violate this principle. People almost invariably make predictions that imply that the relative position of companies or individuals is expected to remain invariant across replications.

The error of nonregressive prediction is as common among experts as among laymen. Indeed, it has been shown that statistical sophistication has little or no effect on the tendency to make nonregressive intuitive estimates. When the subject in an assessment interview states an estimate or a prediction that appears to him as most representative of his view of the situation, the interviewer may safely assume that this estimate is almost certainly too extreme, nonregressive, and therefore nonoptimal.

How is this pitfall to be avoided? How can the subject be guided to apply the information that yielded his initial intuitive estimate in a manner that is more consistent with the principles of statistical prediction? We now outline a series of steps that may be followed to achieve this objective.

- Identify the problem as a member of a broader class for which an average may be assessed, either from past statistics or by relying on the subject's experience. The intuitive estimate will ultimately be regressed toward that average value.

107

- Assess the average and the range of variability for members of the relevant class.

- Use the specific information that distinguishes the particular problem from other members of its class to assess the relative standing of the instance within its class.

- Evaluate the predictive power of this specific information.

- Discuss the regression problem with the subject, if predictive power was judged to be low and the intuitive prediction was relatively extreme, in an attempt to convince him that the best estimate should be regressive when predictability is poor.

We now discuss these stages in turn.

A.5.1.1 Selection of a Relevant Class

Any particular problem can be assigned to many different classes, but it is often easy in practice to select the relevant one. The objective is to relate the specific problem to a class that is as homogeneous as possible and for which an average is known or can be assessed with adequate precision.

Consider again the example of the editor who attempts to predict the sales of a spy thriller set in Russia, written by Mr. X and to be published by Firm Y in a hard-cover edition. It is easy to rule out books about Russia as the most relevant class for the estimation of sales since this class also includes such items as Russian economics textbooks whose sales are determined by entirely different factors. The class of thrillers is not appropriate either because it includes paperbacks as well as hard-cover books. If Mr. X has already published several spy thrillers with Firm Y, the average sale of these books is surely the most relevant value. If this is Mr. X's first book, the class of hard-cover thrillers published by Firm Y may provide a useful substitute. If Firm Y is itself new in the business, the class of initial publishing ventures that introduces hard-cover thrillers may be the most appropriate.

There are two criteria for the selection of a class. Because these criteria do not invariably agree, good judgment is often required to

make an adequate choice. The two criteria are: the confidence with which one can assess an average value for the class; the homogeneity or variability of instances within the class (e.g., the range over which sales vary for "hard-cover thrillers published by the Book-of-the-Week Club").

A modified definition of the quantity to be assessed is sometimes very helpful in defining a relevant class. Consider, for example, an attempt to estimate the actual costs of an R&D project for which a tentative budget proposal has been prepared. Because R&D projects vary widely in their costs, even within a particular technology, the definition of the particular project as a member of the class of R&D projects appears to serve little purpose in the prediction of costs. If the quantity to be assessed is redefined as "the percentage by which actual costs will exceed the initial budget," an appropriate class may soon be found for which much relevant experience exists. It is also possible to retranslate the resulting estimate from the scale of percentages to dollar values after the assessment is completed.

A.5.1.2 Assessments of Average and Variability for the Class

At this point, the subject should be asked to provide estimates of the average and of the range of variability for the selected class. These estimates should be as detailed and documented as possible. Note that the questions refer to the class, not to the individual problem: "How many books of this type are sold, on the average?" "What is the range over which sales vary for books of this type?" The range should be stated in terms of extremely low and extremely high values that still occur within the class.

A.5.1.3 Intuitive Estimation of the Quantity

Some of the information the subject has about the problem will have been absorbed in the definition of the class to which the problem is assigned, but a considerable amount of specific information about the specific case usually remains. The subject may use this specific

information to generate an intuitive estimate of the quantity that locates the particular instance within the range of variability of the class. Subjects normally find it easy to produce such intuitive estimates. As was noted above, however, this intuitive judgment is most likely to be nonregressive, and a more adequate estimate would generally be closer to the class average.

A.5.1.4 Assessment of Predictability

The subject should now be led to a critical evaluation of his ability to predict differences among members of the relevant class on the basis of the type of specific information that is available to him for the specific case. Sample questions are: "If you read two thrillers of this general category, what odds could you give, on the average, in predicting which of the two will sell better?" or, "If you consider two junior executives that you have seen at work for six months, how often would you be right in predicting which of the two will be promoted further in ten years' time?" With little additional probing, people often state that their ability to predict the order of members of a class on the relevant discussion is actually quite limited.

It is important to remind the subject that predictability is affected by the presence of 'noise' or 'chance fluctuations' both in his own information and in the quantity that he is attempting to assess. If the editor's impression of the book could be affected by his mood on that particular day, or if the sales of thrillers depend mainly on erratic fads, the predictive value of the editor's impression cannot be high.

A.5.1.5 Obtaining a Corrected Regressive Prediction

The intuitive estimate should be regressed toward the assessed average of its class when predictability is judged to be moderate or low, and the intuitive estimate is relatively extreme within the assessed range of variability of the class. When these conditions are met, the basic principle of regression toward the mean should be

110

explained to the subject, and he should be informed of the probable direction of bias in his estimate. It is not necessary to enter into a detailed discussion of statistics. Most people will be satisfied with a statement such as "experience with many assessments has shown that the actual value of the quantity is likely to fall somewhere between the top-of-the-head intuitive estimate and the average of the class," or "I should remind you that in uncertain situations things are rarely as good as one hopes or as bad as one fears." There are many sources of uncertainty in this situation, and they are more likely to pull the value toward the average rather than away from it. If the subject has accepted this point, he should be willing to modify his intuitive estimate accordingly.

A.5.2 Anchoring Biases

Another important source of bias in the estimation of quantities is an effect labeled anchoring. Any value that is mentioned to the subject or that he brings to mind himself while trying to assess a quantity will act as an anchor that pulls the estimate in its own direction.

The anchoring effect is best understood as a manifestation of suggestibility. Any value that is mentioned or occurs to one is given some weight in the process of estimation, even when this is quite inappropriate.

To demonstrate the anchoring bias, we asked one group of people to assess the probability that the population of Turkey exceeds 5 million. Another group of subjects was asked to assess the probability that the population of Turkey was more than 65 million. Both groups were then asked to estimate the population of Turkey. The median estimate was 25 million in the former group and 35 million in the latter group. Thus, the mere mention of a value (5 million or 65 million) caused an anchoring effect, although the subjects were explicitly told that the value appearing in the question was selected arbitrarily. Propaganda and smear-tactics actually work on the same principle: it is extremely difficult to ignore any 'information' to which one is exposed, even when

one consciously regards the source of the information as unreliable and unworthy of trust.

Under some conditions, an estimate that the subject provides himself may serve as an anchor in his subsequent judgments of related quantities. This is probably one of the sources of people's difficulties in recognizing the fallibility of their estimates.

Commitments, promises, and deadlines are natural candidates for anchors. A subject who assesses the total cost of a given project, for example, may reason as follows: "Initially, the estimated cost was $35 million. Now, this was not realistic and the actual cost will surely exceed this value. I would guess total cost to be about 45 million." More often than not, this line of reasoning leads to an estimate that remains too close to the original anchor despite the adjustment in the correct direction.

Several recommendations can be drawn from this discussion of anchoring.

(*)The interviewer should carefully avoid supplying anchors in the questions that he formulates. Any number that he brings up is likely to serve as an anchor in the estimation of highly uncertain quantities.

(*)The subject's early 'ball-park estimate' of the quantity will invariably serve as an anchor in his subsequent assessment of the probability distribution. It is therefore important to go slowly and to encourage the subject to explicitly consider the main relevant factors before he commits himself in any way to a particular value.

(*)Anchoring is especially likely to be an important factor in situations of extreme uncertainty, in which the subject has little useful information, or in snap judgments that the subject may make without using all the relevant information actually available to him. The interviewer will often have to accept relatively superficial assessments of some quantities because of limitations of time and patience. When the interviewer has reason to suspect that anchoring on a value has been a major determinant of such a judgment, he should

112

always ask the subject to quickly list other relevant values that could serve as useful anchors. This process takes little time, and it may result in the subject naturally adjusting away from his anchors.

(*)It is helpful to note in this context that most people readily recognize the role of anchoring in their judgments when the effect is described to them. Discussion of the anchoring bias with the subject improves his ability to be critical of his own performance.

A.5.3 Planning Biases

A third source of biases in prediction of values is the common tendency to underestimate quantities such as the time required to complete a project or the total cost of a project. Although these errors can sometimes be attributed to motivational factors, they exist even when the subject has no apparent reason to underestimate cost or time. Even experienced and knowledgeable subjects find themselves repeatedly making unrealistic estimates in the context of planning. This bias is due in part to the chain-like nature of plans. To complete a project as planned, many stages have to be completed on schedule. For example, to finish a building on time, there should be no delays in transportation of materials, no workers' strike, no unusual weather conditions, etc. Because each one of these disturbances is rather unlikely, people discard them altogether and expect the building to be completed on schedule. Although each disturbance alone is improbable, the probability that at least one of them will occur may be substantial, and its occurrence could greatly delay the project and/or increase its cost.

(*)To reduce the planning bias, it is useful to elicit from the subject a list of events (as complete as possible) that might upset the plan. Having listed many such events, the subject may come to realize that while he is incapable of predicting which of them will occur, he should take them into account somehow rather than ignore them altogether.

A.6 The Overconfidence Bias in Probability Distributions

The previous sections were concerned with the assessment of a best estimate for an uncertain quantity. Decision analysis, however, requires the uncertainty about the quantity to be specified in the form of a probability distribution over that quantity.

A probability distribution of a given subject over some quantity cannot be classified as correct or incorrect. It merely summarizes what the subject knows, or better yet what he does not know, about the quantity in question. It is possible, however, to detect and predict systematic biases in such distributions. These biases become evident when the subject produces many probability distributions for uncertain quantities and when the real values of these quantities are matched against his distributions. Two classes of bias may be defined: central tendency biases, and variability biases.

A central tendency bias exists when the true values cluster on one side of the 50th percentile of the distribution. An overestimation bias is present, for example, when the true value is below the 50th percentile for 90% of the quantities. A more subtle type of central tendency bias arises when the subject predicts nonregressively: here, he tends to overestimate quantities that he expects to be high and to underestimate quantities that he expects to be low. In general, the 50th percentile of the subject's distribution will be close to his 'best estimate' of the quantity. To control central tendency biases, therefore, the interviewer should apply the procedures described above in the context of estimation.

A variability bias exists when the subject's knowledge or ignorance about the quantities that he assesses is not properly reflected in the variability of his probability distributions. Suppose we examine many distributions assessed by a subject and record the proportion of cases in which the actual value of the quantity was either smaller than the 10th percentile or larger than the 90th percentile of the respective distribution. Such cases are called surprises. If the subject is properly calibrated, the proportion of surprises should be 20%. If the

114

observed proportion of surprises is much greater, the subject is overconfident; if the percentage of surprises is much lower, the subject is underconfident.

Research shows that experts and laymen alike exhibit considerable overconfidence. In general, the observed surprise rate is about 50% instead of 20%. That is, when people are 80% sure, they are wrong 50% of the time. Two major factors contribute to the overconfidence bias: anchoring and conditionality.

Anchoring occurs because the subject who is to assess the 10th or 90th percentile of his distribution for a quantity typically starts by producing a best guess--i.e., an estimate of central tendency, that serves as an anchor when he turns to the assessment of outlying values. Research has shown that people behave in this manner even when they are not specifically asked to assess their best guess before they assess other values. By an anchoring effect, then, the 10th and 90th percentiles of the distribution will tend to be set too close to the best guess. The resulting distributions will be too tight, thus suggesting a higher degree of confidence than is justified by the subject's knowledge.

A second factor that contributes to overconfidence is the reliance on (unstated) assumptions regarding the assessed quantity. The subject who assesses the revenue of a given company, for example, tends to make a large number of unstated assumptions regarding the company, the market, and the economy. Usually, the subject assumes normal operating conditions and does not take into account the possibility that these conditions might be drastically changed because of war, depression, or sabotage. Indeed, subjects often claim that their expertise is limited to normal conditions and that if these conditions are drastically altered, "all bets are off." Consequently, the subject produces a probability distribution that is conditioned on the assumptions he makes. Such a distribution reflects only part of the subject's uncertainty regarding the quantity, and hence it yields too many surprises. Several steps can be taken to combat overconfidence.

115

(*)When the procedure outlined in the preceding sections is followed with a careful assessment of a best guess preceding the elicitation of the distribution, the interviewer should prevent the subject's initial estimate from being the sole anchor by deliberately suggesting the extremes of the range of variability of the class as alternative anchors. The subject should be asked to consider the possibility that the actual value of the quantity may reach one of these extremes.

(*)An alternative interviewing sequence may be used when the subject is first asked to consider all the factors that could combine to produce extremely high or extremely low values of the quantity and to assess the extremes of his distribution before he is asked to state his best guess.

(*)Whatever sequence is used, it is important to stimulate the subject to think imaginatively of combinations of circumstances that may lead to extreme values. Some of these combinations will appear rather far-fetched to the subject himself and may leave him more convinced than ever that extreme values are quite unlikely to occur. It is legitimate for the interviewer to attempt to shake such convictions by raising questions such as "How sure can one really be that such a constellation will not arise?" or by explicit mention of the overconfidence effect and statements such as "The unexpected does tend to occur more often than one expects."

(*)To reduce the effects of conditionality, the subject should be asked to state explicitly, prior to the elicitation phase, all his presuppositions regarding the assessed quantity. In assessing the revenue of a firm, for example, the subject should stipulate which factors are held constant and which are allowed to vary. The subject may wish to stipulate no major organizational change in the company and no major social, political, or economic changes in the country. The interviewer, in turn, may ask the subject to eliminate some of the assumptions and to incorporate his uncertainty regarding some factors

116

(e.g., public attitude) into the assessments of the quantity under study.

A.7 Judgments of Probabilities for Discrete Events

There are frequent occasions on which a decision analysis requires an assessment of the probability of a discrete event, such as the outcome of an election, the outbreak of a war, or the cancellation of a project. Strictly speaking, a statement of probability is a subjective statement that describes the strength of one's belief that the event in question will occur. Most people, however, interpret a probability statement as an attempt to describe objective reality rather than as a description of what happens in their heads. People who are not trained in modern statistical reasoning spontaneously adopt one of two possible objective interpretations of probability: the probability of an event is the _relative frequency_ with which it occurs, or the probability of an event reflects a _propensity_, the strength of the tendency of a particular system to produce that event. We shall discuss in turn these two interpretations and the biases to which they are liable.

A.7.1 Probability as Relative Frequency

Many probabilities are naturally interpreted as relative frequencies. If one considers the possibility of rain spoiling a planned garden party, for example, the relative frequency of rain at that time of the year is obviously the appropriate value to use. Frequency statistics can be obtained for many problems, and they are preferable to the intuitive judgment of the best of experts. In other situations, however, it is necessary to rely on a subject's assessment of the rate of occurrence of some event.

In general, estimates of rate of occurrence are reasonably accurate for events that are of relatively high frequency _and_ attract some attention whenever they occur. Most people can provide a reasonable estimate of the number of rainy days in a year, but find it quite difficult to estimate the number of times they blink in a minute (unless

117

they spend a few seconds paying attention to their blinks). For events of low frequency, or events that are not noticed when they occur, estimates of rate of occurrence are often inaccurate and subject to major biases. People typically attempt to estimate the frequency of a class of events by the ease with which instances come to mind. This is called judgment by availability. Thus, people assess the probability of a car accident, a heart attack, or a bank failure by recalling instances of these events and assessing the ease with which they come to mind. This mode of judgment is often useful because, in general, instances of frequent events are more available than instances of less frequent events. Nevertheless, it is prone to availability biases.

To illustrate the bias that may arise in judgments of frequency by availability, consider the following question: "Is it more probable that a word sampled at random from typical English text will begin with 'k' or that it will have the letter 'k' in the third position? Most people believe that words that begin with a 'k' are more frequent because it is much easier to call words to mind by their first letter than by their third. In fact, however, 'k' is much more likely to appear in the third position than in the first position in a typical English text.

Because of an availability bias, the frequency of events whose instances are memorable, salient, recent, or dramatic will generally be overestimated, whereas events whose instances are not readily imagined or retrieved will be underestimated.

Uneven coverage of various events in the media is a major source of availability biases, as demonstrated in the public's perception of causes of death. Most people erroneously believe that lung cancer is more dangerous than stomach cancer, that more people die in fires than by drowning, and that homicide is more frequent than suicide. Clearly, lung cancer is mentioned more frequently than stomach cancer in connection with the campaign against smoking, fires are reported much more frequently in the media than drownings, and homicides are widely publicized whereas suicides are rarely mentioned. Thus, the

differential reporting of various events biases their perceived likelihood.

A major source of bias in probability assessment is the impact of the fortuitous availability of incidents or scenarios. Many people have experienced a temporary rise in the subjective probability of an accident after seeing a car overturned by the side of the road. Similarly, many must have noticed an increase in the subjective probability that an accident or malfunction will start a thermonuclear war after seeing a movie (e.g., "Dr. Strangelove") in which such an occurrence was vividly protrayed. Continued preoccupation with an outcome may increase its availability and hence its perceived likelihood. People are preoccupied with highly desirable outcomes such as winning the sweepstakes, or with highly undesirable outcomes such as an airplane crash. Consequently, they are likely to overestimate the probability of these events.

(*)The main way of controlling availability biases is by warning the subject of their possible effect on his assessment. It is sometimes helpful to use another event whose frequency is known as a standard to which the frequency of the critical event is compared. In such a context, the subject may be able to evaluate the degree to which the comparison is distorted by one of the sources of availability bias that the interviewer will have described to him.

A.7.2 Probability and Propensity

Most probabilistic questions of interest are not naturally interpreted in a frequency mode. The probabilities that a defendant is guilty, that Team A will beat Team B in football, that the dollar will be devalued during the next fiscal year, or that there will be an outbreak of hostilities in a given region cannot be interpreted as relative frequencies. In such situations, people develop a model or a schema of the situation and interpret the probabilities of various outcomes as the strength of the propensity of the model to produce each outcome. The stronger the propensity or the disposition of the model to

119

produce an outcome, the more likely that outcome is judged to be. The propensity interpretation of probability provides a useful means for translating knowledge and beliefs regarding a particular situation into judgments of likelihood. For example, if we regard a particular political situation as unstable and explosive, we shall ascribe a high likelihood to the outbreak of a war or some other kind of crisis because our model incorporates the causal dynamics that may generate such outcomes. Indeed, it may be argued that propensities or dispositions are closely related to probabilities. However, there are certain important differences between probabilities and propensities. The common tendency to interpret questions about probabilities as referring to propensities is therefore likely to produce systematic and costly errors.

First, propensities are viewed as characteristics of the system and as such are not very sensitive to time. For example, the propensity of an explosive political situation to lead to war is not very sensitive to the duration of the time frame we consider. The probability of an event, on the other hand, increases consistently with the width of the time frame. Thus, a propensity interpretation of probability may lead people to overestimate the probability of an event occurring during a particular month, and to underestimate the probability of the same event occurring during a specified year.

A second feature of propensity judgments is that these judgments are generally insensitive to considerations of unreliability and predictive accuracy. For example, we are likely to attribute to a defendant a propensity for violent action if we are told that he is "quick-tempered, hot-headed." The reliability and/or validity of that verbal report does not appear to have such impact on our presumption of the defendant's propensity to commit a violent act. However, the probability that the defendant has in fact committed a particular act of violence depends critically on the reliability of that report. An unreliable report should have no impact on judged probability, but there is much evidence showing that information that can be translated into

120

propensity is used even when it is admittedly scanty, unreliable, and invalid. Thus, a major source of bias resulting from the assessment of probabilities by propensity is a lack of sensitivity to the reliability of the input information.

The reader will recognize the similarity of this common error in the judgment of probability to the error of nonregressive prediction of values that was described in an earlier section. Indeed, similar mental activities are involved in predicting a value that is most representative of one's model and in judging the probability of an outcome by the propensity of that model to produce the outcome. The subject merely uses the scale from 0 to 100 to express his impression of propensity, but the process is the same as when he uses a scale of dollars to express an intuitive impression of the future earnings of a firm.

Some of the corrective procedures that may be applied to judgments are also similar in the cases of probability assessments and prediction of values.

(*)Whenever possible, the event whose probability is to be assessed should be viewed as a member of a larger class of events to which it is similar in some essential respect and for which a rate of frequency of occurrence can be obtained from statistics or estimated with reasonable precision. The specific features that distinguish the particular problem from other members of the class should be used to adjust the estimate above or below the value for the class. The extent of adjustment should be controlled by an assessment of the predictive value of the specific information.

Consider, for example, the following problem:

> A man has been drawn at random from the adult population. This man has been described by a casual acquaintance as "meek, tidy, and with a passion for detail." What odds would you give that this man is a librarian rather than a farmer?

The immediate intuitive answer of most people is that the man in question is very likely to be a librarian. The procedure described above, however, would soon draw attention to the class from which the

case has been drawn. This class is the adult population, where there are many more farmers than librarians. Adjusting the estimate from the relative frequency of librarians will yield a much lower probability value than the untrained intuitive guess. Consideration of the reliability and validity of the information may cause one to further moderate the extent of the adjustment.

This procedure is not always applicable, however. There are many problems (e.g., the likelihood of a war between Greece and Turkey) which cannot readily be assigned to a meaningful class for which a frequency can be assessed. An individual who assesses this probability is most likely to do so by a judgment of propensity, but the interviewer may have little to contribute beyond the obvious suggestion that the respondent should consider all the available evidence. The obtained estimate, however, should be treated with extreme caution. Intuitive assessments of the probability of truly unique events cannot be corrected, and uncorrected assessments of propensity are subject to massive errors and biases.

A.8   The Interpretation of Conditional Probabilities

Many of the assessments that are required for decision analysis refer to conditional probabilities or to conditional probability distributions (e.g., the probability that a particular project will receive continued support if Mr. X is elected President or the distribution of future development costs for a project if a competing firm markets its version of the product by a specified date).

In principle, an assessment of probability is always conditional, because the assessment implicitly assumes everything that the subject believes to be true about the world. The difference between standard and conditional probabilities is that, in the latter case, the subject is required to assume a present or future state of the world that differs in a specified way from his current model.

The proper interpretation of a conditional probability is that the statement of the condition provides additional information about the

system and thus requires an appropriate modification of the entire framework within which the judgment is to be made. Consider the example of an assessment that is conditional on the outcome of a presidential election that is due to take place in a year's time. The subject will draw his assessment from a complex image of the entire system, including the local conditions, a view of the general economic situation, and some idea of the tendencies of these conditions to change in various ways during the coming year. When given the added information that Mr. X will be elected, rather than Mr. Y, the subject should use this information to modify all aspects of his general model as in the following example: "If Mr. X is elected, this must mean that the economic situation .at the •time is likely to have been..." The assumption of the condition alters the probabilities of the target event by changing the most probable _initial_ state of the entire system.

People do not normally evaluate conditional probabilities in this manner. They do not modify their model of the situation, but merely add an impression of the causal effect of the condition to their current impression of the propensity of the system to produce the target event. To return to our example, financial support for the project may depend both on the general state of the economy and on the president's policy. Most subjects will alter their assessment of probability (relative to the unconditioned case) only by considering the impact of Mr. X's policy, without altering their view of probable economic conditions to fit the fact that Mr. X will have been elected.

This failure to properly use the information conveyed by the condition can lead to major errors and inconsistencies in the assessment of conditional probabilities. An example recently published in a forecasting journal illustrates problems. Consider the following:

(A) By January 1980, more than 500 cases of death will have occurred that will be attribued to mercury poisoning.

(B) By January 1980, Congress will have passed a law limiting mercury pollution.

Most people who are required to make assessments assign values such that:

$$P(A|B) < P(A|\bar{B}) \ ,$$

where $\bar{B}$ refers to the negation of B--i.e., to the event that no such law will be passed. The same assessors also assign values such that:

$$P(B|A) > P(\bar{B}|A) \ .$$

The reader may find it useful to ponder these assessments, which initially appear very compelling.

The trouble is that these plausible judgments violate the most elementary principles of probability theory, where the statement

$$P(A|B) < P(A|\bar{B}) \text{ implies that } P(\bar{B}|A) > P(B|A) \ .$$

How do these inconsistencies come about? The reason is that, for any given model of the world, the passsing of a pollution control law does reduce the likelihood of subsequent disasters. For any given state of the world, the occurrence of a pollution disaster increases the propensity of Congress to pass a pollution control law. Subjects evidently base their estimates of conditional probabilities solely on the presence of these obvious causal relationships. They fail to consider the statement of the condition as a source of information and view it entirely as a causal agent. Upon reflection, it becomes evident that the assumption that a pollution control law has been passed provides some information about the conditions that preceded this event. Specifically, it suggests that a major disaster could very well have occurred, that then prodded Congress into action. In this manner, the information conveyed by the stated condition can be used to alter one's probabilistic model of the initial state of the system. This is precisely what people normally fail to do.

To overcome this critical flaw in intuitive assessments of conditional probabilities, the following procedure is recommended.

(*)When a conditional assessment of probability is required, the subject should be encouraged to speculate about the various

124

circumstances that could cause the condition/event to occur. This is especially important when the condition/event is unlikely, in that the required modification of the overall model is most profound in such cases.

By speculating about the explanation of the condition, the subject will be led to a view of a probable state of the system, given that the condition/event occurs, which may be different in many respects from his current model. This phase of the discussion should be sufficiently prolonged to allow the subject to form a fairly complete and coherent alternative model. In subsequent assessments of conditional probabilities, the subject should assume his alternative model of the situation in its entirety and not merely that a single condition/event has occurred.

Appendix B

PROCEDURE AND RATIONALE FOR THE PROBABILITY ENCODING INTERVIEW

The encoding interview is viewed as a joint undertaking of the subject (the interviewee, the expert) and the interviewer (the analyst). The purpose is to provide a record of the subject's probability distribution for a particular uncertain quantity (hereafter referred to as _the_ Quantity), and an explicit and comprehensive record of the considerations underlying the subject's judgments. The interview form is designed to aid the interviewer in eliciting the subject's considerations, in anticipating and reviewing likely biases, and in providing a convenient form for the report of the encoding interview.

Encoding sessions vary widely because of differences in the Quantity and in the people participating. It is therefore not possible (or at least not meaningful) to design a set of standard questions to be used in all encoding sessions. We have instead chosen to illustrate each step with one or more sample questions to give a feeling for how questions may be formulated. In some cases, a question can be used directly; in other cases, the interviewer will have to design his own questions.

The procedure is quite long the way it is presented here. However, this way it covers many of the situations that may come up in an interview. The interviewer should feel free to skip certain parts if he does not find them applicable.

This appendix is essentially a self-contained document--i.e., an interviewer can take it, study it, and apply it in an interview. However, we believe that it will be easier to understand the procedure if the interviewer has read the Manual first.

I MOTIVATING

The first phase of the encoding procedure has two purposes. One is to establish rapport with the subject. The other is to explore whether the subject's responses might become biased because of his perceived system of personal rewards. The motivating stage is essential in that it encourages the subject to become actively engaged in the encoding process and to view his responses as being important to the solution of the decision problem. The encoding process can be time-consuming, and the subject must feel that what he is doing is necessary and worth his attention.

Step 1: Introduction to the Encoding Task.

The introduction may entail an explanation of the purpose of probability encoding in decision analysis. An explanation of what decision analysis is expected to contribute to the actual decision problem may also be included. The interviewer might mention (or ask) why the subject has been selected as a contributor of judgment regarding the Quantity to be discussed. The importance of the Quantity to the decision problem may also be pursued with the subject to give him a sense of why time is being spent encoding the Quantity.

The interviewer should discuss, if necessary, the difference between deterministic (single number) and probabilistic (probability distribution) forecasts. It should be made clear to the subject that the encoding is not concerned with predicting the outcome of the Quantity, but rather with generating a description of the overall uncertainty that the subject feels regarding that outcome.

Step 2: Exploration of Motivational Biases

The interviewer and the subject should discuss openly any personal payoffs that may be associated with the probability assignment and the ramifications of possible misuses of the information. The subject may

be aware of misuses of single-number predictions--e.g, that they often are interpreted as firm projections or commitments. It should be pointed out that no commitment is inherent in a probability distribution because the distribution shows that there is a range of possible outcomes. In fact, the only aim of the encoding process is to develop a probability distribution that represents as clearly as possible the complete judgment of the subject.

If the subject is involved in some way with the Quantity (he may be a project manager who is asked about the completion time of his project or a sales manager who is asked about next year's sales of a particular product), the discussion should be directed to factors that may affect the outcome of the Quantity but that are outside the subject's control. It should be explained that the subject cannot be held responsible for every aspect of the outcome. This may help him provide unbiased judgment.

During the discussion, the subject is likely to reveal some reactions indicating biases that might be expected later in the encoding session. The interviewer can discuss potential biases with the subject and try to lead him toward providing complete and open judgments concerning the Quantity.

Extensive note-taking can be a useful device to avoid a motivational bias. Because these notes will remain on record and be open to general review, the subject would be encouraged to give a balanced presentation of arguments. In a way, the arguments and considerations are as important as the actual numbers encoded.

The following are examples of representative interview responses indicating motivational biases.

Example 1: Total Cost of a New Weapon System

> Question: Joe, you've been selected to provide information concerning the total costs of the new weapon system. Why were you chosen?

130

Answer: Well, I suppose I am the one who is going to be the program leader for the system. That job will be a real opportunity for me to advance. I should be able to jump a couple of grades on this assignment alone.

Remark: The subject may want to influence the decision and therefore expresses his judgment in a way that appears more favorable to one decision alternative. For example, since he will be the program leader, he may want to ensure that the program is continued. He might view cancellation of the program as a personal loss to his career.

The interviewer can try and emphasize that the analysis and its *recommendations* will consider the whole range of potential outcomes for the Quantity and weigh these outcomes by their probabilities of occurrence. In this way, the interviewer can encourage the subject to think about both favorable and unfavorable outcomes. The interviewer can also make the subject aware that being program leader on a program that is causing problems and experiencing budget overruns may not be conducive to career advancement. If in the end it becomes clear that the subject is biasing his responses, the interviewer may try to restructure the questions so that the subject does not know how to answer to best serve his interests. Finally, if this tactic does not work, there may be no way out but to disqualify the subject.

Example 2: Program Cost

Question: In the past, you have had to make up forecasts of costs for your programs. How have your forecasts compared with actual cost figures?

Answer: Always high. I purposely forecast high. That way I come in under budget, which is good at review and promotion time.

Remark: A program manager may consciously give a high prediction of program costs because he thinks he will look good if the actual amount is less than his forecasts, or because there are punitive incentives in the system for anyone who underestimates.

131

The interviewer can tell the subject that there is *no right* or wrong answer with a probabilistic forecast--sometimes the actual outcome will be in the low end of the distribution, sometimes in the high end. Stress that there is no commitment in a probabilistic forecast and that the information will not be used to measure his performance. If necessary, the decision maker may have to be called in to confirm this assertion.

Example 3: Field Lifetime of a New Weapon System

> Question: You have been selected to provide probabilistic information for the field lifetime of the new system. Do you feel like you can provide data on this subject?

> Answer: Yes, I've been doing exclusive testing of the prototype system for the past year. In fact, I've become the resident expert. I can probably give you estimates of lifetime for that new system within plus or minus 5%.

> Remark: The subject may feel that because he is an expert on the subject matter, his range of uncertainty should be narrow. In other words, he thinks he is expected to know the answer.

The interviewer can explain that the subject *is not expected to* predict the exact outcome and that the range of uncertainty can some-times vary widely, even for people who are recognized as the most knowledgeable. An example that might be presented is that of a pilot giving a distribution of a plane's effective radius of operation. The pilot, although an expert, is very unlikely to give a narrow distri-bution for the Quantity once he takes into account all the possible factors such as speed, altitude, wind activity, and so forth, that may influence the plane's flight activities. The implications of outcomes falling outside narrow distributions can also be discussed with the subject.

A training session would be useful if time permitted. It would *give the subject a* better understanding of the relationship between

knowledge and uncertainty. Most subjects make few probability assignments in the course of their work; therefore, they do not receive extensive feedback about past performance. It is then possible to provide training through experimental sessions in which the subjects are asked about any kind of uncertain quantities. Because we want to improve their ability to quantify judgment, it is immaterial whether we use almanac questions (such as "What was the number of automobiles registered in the U.S. in 1977?") or quantities relating to their field of expertise.

II STRUCTURING

This phase has two purposes. One is to define and clearly struc-
ture the Quantity. The other is to explore how the subject thinks about
the Quantity, what information he might use, how he relates the Quantity
to other variables, and so on. This phase also serves to decide whether
further modeling is necessary and whether the assumptions for the
encoding should be modified. The record of this phase of the interview
should be clear and complete so that it may be evaluated by other
experts or by the decision maker. This will motivate the subject to
consider all aspects of the problem and to base his judgments on defen-
sible considerations. The structuring phase is analogous to the deter-
ministic phase of decision analysis, and it may reveal the need for
further modeling, redefining the Quantity, or making the Quantity
conditional on some other variable.

Step 1: Definition of the Quantity

Here the interviewer defines the Quantity and states the conditions
that are assumed to hold. He verifies that the definition and the
conditions meet the clairvoyance test--i.e., that a clairvoyant could
reveal the value of the Quantity by specifying a single number without
requesting clarification. For example, it is not meaningful to ask for
"the price of wheat in 1980," whereas "the closing price of 10,000
bushels of durum wheat on June 30, 1980 at the Chicago Commodity
Exchange" is a well-defined quantity.

Step 2: Approaches to the Assessment of the Quantity

The subject is asked to describe possible approaches to the
assessment of the Quantity and to list the variables that should be
considered according to each of these approaches.

135

Example: The Cost of a Program for a Given Year

Question: How would you break down or reformulate the problem?

Answer: (a) Break down source of costs by department.

(b) Break down costs in terms of manpower, utilization factor, costs per unit time, and equipment charges.

(c) Reformulate problem in terms of past costs, program expansion, and inflation factors.

Remark: At this stage, a decision may be made to model further, to separately encode quantities mentioned by the subject (e.g., costs by each department), or to change the assessed Quantity (e.g., encode rate of growth of cost instead of actual costs). The variables and units with which the subject is most comfortable should be selected.

Step 3: List of Relevant Factors and Information

The interviewer elicits factors or scenarios that could affect the Quantity. The purpose is to encourage the subject to hold in mind a comprehensive model of the uncertain situation and of the extreme outcomes to which some combination of factors may lead. The interviewer should note the subject's answers and repeatedly probe to obtain a comprehensive list, stating the incompleteness principle (below) if appropriate.

Incompleteness principle: When people say they have considered everything they could think of, they usually haven't.

Example: Budget of a Program

Question: What factors will have a major effect on the budget? Under which circumstances could the budget be high or low?

Answer: General state of the economy; election-year activities; Presidential policy; test results of prototype systems.

Further probing: What else? Can you think of other determinants? Let me remind you that when we feel we have covered all factors, we usually have omitted a few important ones. Try to think of some more.

Further   World military activity; attitude of potential
answer: users of this system when they preview the
              system next month.

Step 4: List of Assumptions

The interviewer elicits assumptions that the subject is making in thinking about the Quantity. For each factor that is mentioned in this step, the interviewer and the subject should check whether it is included in the definition of the problem. If it is not, then the initial list of conditions must be expanded, or the subject must be instructed to consider the factor as a variable.

Example: Future Budget of a Program

Question: In thinking about the budget for next year, what assumptions are you making? Which things do you think will not change?

Answer: The key people will remain on the project. No major organizational changes.

Probing: What else won't change between now and the end of next year?

Further   No major social, political, or economic
answer: changes.

Summary   Let us stipulate that there will be no major
comment: organizational changes. All other factors should be viewed as uncertain, and you should consider different possible values of these factors in your assessment. This includes personnel changes on the program and all political or economic changes.

Step 5: Redefinition of Problem

In light of the structuring phase, the interviewer and subject review the definition of the Quantity and write down all the conditions assumed for the encoding. Time is taken to develop specific descriptions of the Quantity that satisfy both subject and interviewer and that seem to eliminate as much ambiguity as possible.

Step 6: Choice of Scale

Use a scale for the Quantity that is meaningful to the subject. The wrong choice of scale may cause the subject to spend more effort fitting his answers to the scale than evaluating his uncertainty. After the encoding, the scale can be changed to fit the analysis.

Example 1: Amount of Oil Required per Year To Heat a New Plant

Question: How would you measure the amount of oil? In tons, gallons, barrels, tank cars, or what?

Answer: I normally think in terms of tons.

Example 2: The Value of the Swiss Franc in U.S. Dollars on December 31, 1979.

Question: When talking about the value of the Swiss franc, can we use the value in dollars or cents of one Swiss franc?

Answer: You know, I am more used to thinking about the value of francs per dollar.

Comment:* Fine. Let us do it that way. We should use whatever is most convenient for you. [It is a simple matter to change the encoded distribution to the corresponding one for the Swiss franc expressed in dollars.]

Example 3: Number of People Needed by the Program Next Year

Question: How do you want to think about the number of people next year? Do you want to think about the actual number, or do you prefer to think in terms of the change from this year's level of staffing?

_____

*"Comment" represents the interviewer's comment to the subject.

Answer: I would calculate next year's number by increasing this year's number.

Comment: So the variable of interest is really the level of increase from this year to the next. We will then concentrate on that increase.

III   CONDITIONING

The purpose of this phase of the interview is to draw out the subject's knowledge relating to the Quantity so as to give him a conscious basis for making probability judgments and to counteract encoding biases that he might otherwise exhibit. The phase consists of two blocks. The first block is aimed at defining the information on which the subject is basing his judgment and at achieving a proper balance between general and specific information. The second block is aimed at counteracting anchors but will also compensate for availability and unstated assumptions.

This phase takes the form of a preliminary discussion of values that are informally elicited from the subject. These values should be discussed in the light of psychological and statistical considerations so that the subject may revise them if he finds it appropriate.

As a minimum, this phase should cover Step 9 (in Section III.2). This step provides a general compensation for anchoring, availability, and unstated assumptions even when these biases have not been formally detected.

III.1   Balancing General and Specific Information

Step 1:   General and Specific Information

The information on which the subject will base his judgment of the Quantity will generally have been brought out in the discussions in the first two phases (in particular in Step 3 of the Structuring phase). If the interviewer believes that there is more information underlying the subject's judgment, then it should be educed now.

Next, the interviewer should judge whether the information is to be characterized as general, specific, or of both kinds. General information is information about quantities similar to the Quantity, whereas specific information pertains specifically to the Quantity. For example, consider the annual maintenance cost for a new type of

aircraft. Specific information would refer to the design of the aircraft, the materials used for various parts, and the potential use of the aircraft. General information, on the other hand, would concern maintenance experience with similar aircraft but possibly of other sizes, designs, or applications.

Interviewer's judgment: If the information can be characterized as either all general or all specific, then there is no problem to be solved in this area and the interview should advance to Section III.2.

Step 2: Revision Based on Recent Information

This step applies only if the specific information is restricted to a recent piece of information, e.g., an intelligence report or a field study. Otherwise, the interview should continue with Step 3.

In this step, the subject is first asked to think back on what estimate he would have made prior to receiving the specific information. He should then consciously consider the specific information, its validity and reliability, and what its impact should be. Finally, he is asked for a revised estimate of the Quantity that includes all of his knowledge, both general and specific.

Example: Number of Orange Aircraft in Province X

Question: You mentioned that you received an intelligence report last week that indicated a major buildup of Orange activities in Province X. I would like you to go back a week and recall what your thinking was before you received that report. That is, assume that you had never received the information. What would then have been your top-of-the-head estimate of Orange aircraft in Province X?

Answer: I'd say around 60 aircraft.

Question: Now let's consider the intelligence report. How relevant is it to the problem of estimating the number of Orange aircraft?

Answer: I think it's directly relevant since it is
              concerned with activities that indicate a
              buildup of Orange forces. Surely these will
              also include aircraft.

Question: I agree that the information seems relevant.
              But, if you think back on other times that you
              have received similar intelligence reports and
              on what subsequently happened, would you say
              that these reports represent reliable
              information?

Answer: I see what you mean. In many cases, the
              situation never becomes as serious as the
              intelligence might indicate. But on the other
              hand, one can't disregard the information.
              After all, a number of high officers have been
              seen in the area.

Question: Now, if you combine your general background
              information with that provided by the
              intelligence report and keep in mind the
              reliability of such reports in general, what
              would be your estimate of Orange aircraft?

Answer: You can't forego the report completely, but I
              agree that *we shouldn't jump to any conclusions.*
              How about making it 80 aircraft?

Remark: The interviewer should now proceed to Section III.2.

Step 3: Definition of a Reference Class

Example: Reliability of a New Machine

Question: When you are studying an uncertain quantity, it
              often helps to think about it as a member of a
              broader class of problems that are similar in
              nature. For example, when considering the sales
              of a new book, it can be related to the sales of
              books by the same author or with a similar
              subject matter. If you are estimating the cost
              of some development program, you may think of
              the deviation of cost from budget for other
              development programs. We are now talking about
              the reliability of the Theta machine. Can you
              think of such a reference class in connection
              with Theta?

143

Answer: I can relate it to similar machines that we have developed in recent years.

Comment: So we will use the reliability of similar machines as a reference class.

If the subject cannot define a reference class for the Quantity, the interview will continue with Step 4A, which is an alternative to Step 4.

Step 4: Assessment of the Average and the Variability for the Reference Class

Note that the questions refer to the reference class, not to the individual Quantity.

Example 1: Reliability of a New Machine

Question: What is the average time between failures for machines of this kind?

Answer: About 150 hours between failures.

Question: And how can that number vary among different types of machines?

Answer: I'd say the range is from 90 to 300 hours.

Example 2: Cost of Development Program

Question: How do program costs for these kinds of development programs compare with budget, on the average? How much can program cost vary relative to budget for programs of this kind?

Answer: Oh, they can vary a great deal, say from 80% of budget to maybe 250%. An average might be 15% above budget.

Remark: The interview continues with Step 5.

Step 4A: Less Informed Estimate

This step replaces Step 4 if no reference class could be defined in Step 3. The purpose is to reach an estimate based only on general

144

information so that the estimate can be used instead of the average value of the reference class.

The subject is asked to guess what estimate of the Quantity would be given by another subject competent in the same general area of knowledge but who does not have access to the detailed information concerning the problem. In formulating this question, the interview should define the 'hypothetical subject' so that the subject has the minimal amount of specific information that will permit him to make a meaningful estimate.

Example: The Number of Tank Units in Use Within Five Years

Question: Consider a test engineer like yourself who is told that a tank is being produced that will be able to work well in both desert and tropical environments. What would be his guess about the number of such tanks in use within 5 years?

Answer: If that's the only information he would have, I'd say around 4,000 tanks.

Remark: The interview now continues with Step 5.

Step 5: Intuitive Estimate of the Quantity

The subject is asked to produce a top-of-the-head estimate of the Quantity that best represents his intuitive impression. He should be encouraged to produce an intuitive judgment that will be treated as a basis for further discussion and not as a prediction to which he will be committed. In the same spirit, the subject will be asked to describe how this value occurred to him, but not to defend it.

Example 1: Number of Orange Troops

Question: If you were to give a top-of-the-head estimate of the number of Orange troops that would represent your current impression, without too much deliberation, what value would you use?

Answer: 33,000.

145

Question: How did this value occur to you? What made you
             think of this value?

Answer: The last time Orange performed this exercise,
             they used 30,000 troops. I think this one is at
             a little higher strength.

Example 2: Cruising Speed of a New Ship

Question: How fast a ship do you think this will be?

Answer: Most likely above average.

Question: What does "most likely above average" mean in
             terms of cruising speed?

Answer: I'd say around 25 knots.

Example 3: Unit Manufacturing Cost

Question: You stated $3,000 as a representative value.
             How did you arrive at this value?

Answer: We've been told to keep the cost under $2,500.
             I don't think this is realistic, but we can come
             close.

Example 4: Unit Manufacturing Cost

Question: How did this value occur to you?

Answer: It just looks like a reasonable value.

Question: Could you say more?

Answer: We've been doing pretty well on that problem,
             and if we continue, that's the figure we should
             arrive at.

Question: Are there any plans, standards, or values that
             people talk about a lot?

Answer: We know that it can't be less than $2,500, nor
             more than $4,500.

Remark: The subject's answers in Examples 1 and 3 suggest
that he may have reached his estimate by adjusting from a

146

salient value or from a plan. This will be discussed further in Section III.2.

Step 6: Evaluation of Predictability

Interviewer's judgment: If the subject's top-of-the-head estimate is reasonably close to the average value of a reference class--that is, much closer to the average than to the extreme values of the class--this means that there cannot be an overly high emphasis on specific information. The interview can then go directly to Section III.2. The interviewer should be aware that one reason the estimate is close to the average value of the reference class could be that the estimate was obtained by an adjustment away from the average value, which thus served as an anchor. This potential bias will be counteracted in Section III.2.

The subject is asked to evaluate how easy or difficult it is to predict the outcome of the Quantity on the basis of the specific information available. He should be reminded that the predictability depends on the variability of the specific information and on the validity (the relevance) of the specific information.

Example: Development Cost

Question: If you were considering two development projects in this general category, how often would you be able to predict which of the two projects would have the lower development cost, based on specific information of the kind that you have for the present project?

Answer: Well, if the projects are well defined, it should not be too hard to tell which would be the least expensive. Maybe four times out of five.

Alternative   For some programs in this general category, the
question: cost can be known rather well even two years ahead of time. For other programs, there is a great deal of uncertainty regarding development costs. Where would you place the present program along that scale?

147

Answer:	I think we know enough about this project to be able to predict the cost fairly well.
Alternative question:	As you know, some quantities, such as the attendance at the football games of one team next year, are highly predictable. Other quantities, such as the total audience for a new Broadway play, are highly uncertain. At the present point in time, how predictable do you consider the development cost on a scale from highly predictable to highly uncertain?
Answer:	It isn't highly predictable but we do know quite a bit about the project at this stage.

Interviewer's judgment: If predictability is high, then the interview can skip to Section III.2.

Step 7: Discussion of Extremeness of Intuitive Estimate

This step applies when the intuitive estimate is relatively extreme and predictability is moderate or low. In all other cases, the interview would skip to Section III.2.

The subject should be informed that an intuitive estimate that is relatively extreme within the assessed range of variability of the reference class is likely to be biased when predictability is low. When appropriate, the interviewer should bring up the moderation principle. The subject should then be asked whether he would like to modify his estimate.

Moderation principle: "When predictability is limited, things are rarely as good as one hopes nor as bad as one fears."

Example: Reliability of a New Machine

Comment:	Your top-of-the-head estimate was 250 hours between failures, which seems closer to the high end of the reference class than to the average. At the same time, it seems as if the reliability is still quite uncertain because of limited large-scale test results. I should tell you that experience with such uncertain quantities has shown that the actual value of the quantity is likely to fall somewhere between the average value of the reference class and the

148

> top-of-the-head estimate. In this case, it seems that the various uncertainties are more likely to reduce the number of hours between failures, since reliability can hardly be higher but there is room for it to be lower.

Question: In light of this comment, would you like to revise your top-of-the-head estimate of the number of hours between failures?

Answer: Well, I still believe it's going to be better than average, because of the new design with fewer moving parts, but you may be right about all the things that can go wrong. How about 200 for a rough number?

In case no reference class could be found in Step 3, this step applies when the intuitive estimate is substantially different from the less informed estimate in Step 4A and predictability is moderate or low. The question will have to be phrased slightly differently.

Example: Number of Tank Units in Use Within Five Years

Comment: Your top-of-the-head estimate for the number of tank units was 2,000 units. You also mentioned that another test engineer who had heard about the tank but was not privy to field test data would give an estimate of 4,000 units. You also said that the number of tank units is very unpredictable. I should tell you that experience with such quantities has shown that the actual value of the quantity is likely to fall somewhere between the estimate of the more informed expert and the estimate of the less informed expert. By this rule, one would say that the number of tank units is more likely to fall above 2,000 than below it.

Question: Based on this, would you want to revise your top-of-the-head estimate?

Answer: Maybe 2,000 units is too low considering that the field test results are not conclusive, even though they do not look too encouraging. I'll settle for 3,000 units.

III.2 Counteracting Anchoring, Availability, and Unstated Assumptions

Step 8. Discussion of Anchoring

When the subject's answers suggest that he may have reached his intuitive estimate by adjusting from a salient value or from a plan (see Examples 1 and 3 in Step 5), there is reason to believe that the adjustment is insufficient--i.e., that the judgment is anchored. In such cases, the interviewer should introduce the adjustment principle. If the subject becomes convinced that his judgment was subject to an anchoring bias, he should note the probable direction of bias in his estimate. The discussion in Step 9 of scenarios for extreme outcomes of the Quantity will also help counterbalance an anchoring bias.

> Adjustment principle: When you adjust a value you rarely adjust enough.

Example 1: Number of Orange Troops

Comment: You mentioned that you reached the value of 33,000 troops by adjusting from the number of troops at their last exercise and since you believe that they are at a higher strength this time. I should mention that experience with judgment of this kind has shown when people reach an estimate by adjusting from a value, they typically do not go far enough in adjusting. In view of these results, one could guess that an improved estimate might be even higher than 33,000.

Example 2: Unit Manufacturing Cost

Comment: You said you arrived at the figure of $3,000 for the manufacturing cost by correcting the unrealistic target of $2,500 that management had set. Let me point out to you that when people make estimates by adjusting from a plan or a target, they typically remain too close to the initial value. This is sometimes called the planning fallacy. You may want to consider whether your own top-of-the-head estimate was overly optimistic and whether a realistic estimate of the manufacturing cost would be somewhat higher.

150

Alternative comment:	You said you arrived at the figure of $3,000 for the manufacturing cost by starting from management's target of $2,500. Are there any other numbers that come easily to your mind?
Answer:	Well, the previous model had a cost of $3,700 and we even had one before that which cost close to $4,500.

Step 9: Discussion of Scenarios for Extreme Outcomes

The purpose of this step is to bring to the surface more of the subject's knowledge on which he may base his judgment. On the basis of the previous discussion, the interviewer proposes some very extreme values to the subject. Alternatively, the subject is asked to generate these values himself. The subject is then asked to regard the situation retrospectively from the future. That is, he assumes that he is told at some future time that such an extreme value had occurred, and he is asked to describe a scenario that would explain this outcome. This will help compensate for anchoring and availability and may also bring up unstated assumptions that the subject may have been making.

The step includes discussion of scenarios for both a high value and a low value. The examples below are given for one extreme only.

Example 1: Total Sales of a New Product

Question:	Give me a number such that you would be very surprised if the total number of units sold turned out to be greater than that number.
Answer:	I'd say around 12,000.
Question:	This value is less than the high value for the reference class. Is there any reason?
Answer:	I guess not. Change the number to 15,000.
Question:	Suppose I told you for a fact that more than 15,000 units were sold. Could you give me a scenario that is consistent with this?

151

Answer: Its performance could be at the high end of what
             is now considered possible. Our European
             competitor may drop out of this line. The
             number could expand if many new uses are found
             for the product. Finally, the production cost
             may be lower than expected, which would allow an
             introduction even on small markets.

Question: Considering this scenario and any others that
             could lead to a value greater than 15,000, what
             is roughly the probability of finding sales
             greater than 15,000 units? Is it 1 in 100, 1 in
             10, 1 in 5, or what?

Answer: I'd say around 5% or 10%.

Example 2:   Number of Orange Aircraft in Province X

Question: Suppose I told you for a fact that the number of
             Orange aircraft actually exceeds 200. Could you
             give me a scenario that would be consistent with
             this?

Answer: That's an awfully high number, you know.

Question: Try all the same. Could there be any ex-
             planation of the number of aircraft being at 200
             or above?

Answer: Well, it would mean that the whole air force has
             been moved to Province X, which in turn
             indicates a potential attack on the Green
             Republic. However, that's not likely because of
             the peace treaty signed last year. Another
             explanation may be that Red has increased its
             supply of aircraft to Orange. That's a
             possibility that I haven't considered; it's
             outside of this problem, isn't it?

Comment: Let's then be explicit on this point. When you
             consider the number of Orange aircraft, we will
             assume that there haven't been any further
             supplies by Red. We can discuss that factor
             separately. However, I would like you to
             include the possibility of hostile action when
             you make your estimates regarding the number of
             aircraft.

152

Question: With this understanding, what is the probability
              of the number of aircraft being greater than
              200?

Answer: Maybe 1 in 20 or even less than that.

Remark: The second example brought up two ways to deal with
unstated assumptions. For one case, the subject was asked to
consider the factor when forming his judgment. In the other
case, the assumption was included as a clearly stated
assumption.

IV ENCODING

The purpose of this phase of the interview process is to quantify the subject's judgment in terms of a probability distribution.

> Plotting: All values or probabilities elicited in this phase are plotted on a probability graph paper outside of the subject's view. Begin by plotting the high and low estimates (given as answers to Step 9 in the Conditioning phase). Choose a scale for the plotting that is easy to work with and that makes good use of the width of the graph paper. The ideal scale would have each of the high and low estimates plotted 1 to 1-1/2 inches from the ends of the scale.

Step 1: Elicitation of Probabilities

The interviewer uses the probability wheel to encode probabilities corresponding to different values. Begin by taking a value that is not expected to be extreme (say around 30% of the distance from the low to the high estimate) and encode the corresponding probability level. Make the first choice easy for the subject by making the orange sector on the probability wheel much smaller than what might actually correspond to the subject's probability. Next, choose a sector that is much too large. Then try to find the indifference probability.

Example: Unit Manufacturing Cost for a New Machine

Question: Consider the following two bets. With one bet we can spin this wheel and if it stops with the pointer in the orange sector you win $100; otherwise, you win nothing. [The wheel is set at 10% orange and 90% blue.] With the other bet, you win $100 if the manufacturing cost is below $4,500. Which bet do you prefer?

Answer: I will take the cost being below $4,500; that is more likely than the wheel stopping on orange.

Question: So we have to increase the amount of orange. [The wheel is set at 85% orange and 15% blue.] Again, you can either bet on the wheel coming up orange or the cost being less than $4,500.

155

Answer: Now I would take the wheel.

Question: How about this amount of orange? [The wheel is set
            at 25% orange.]

Answer: I get the point. Reduce the orange a little and I
            will be indifferent.

Question: How about this? [The wheel is set at 20% orange.]

Answer: That is about right.

Remark: The probability of the manufacturing cost being below
$4,500 is thus assigned a probability of 0.20. The response
is plotted on the probability graph paper.

The interviewer continues to elicit probabilities, moving from one value to another without a pattern. Ask for cumulative probability levels (the fixed event is defined as the Quantity being less than or equal to a given value) or their complements (the fixed event is defined as the Quantity being greater than a given value). Elicit around five to ten probabilities.

The probability wheel makes it easier for the subject to express his judgment than using direct questions (e.g., "what is the probability that the cost will be below $7,000?"). However, the latter may be used if the subject does not understand the procedure with the wheel or if he finds it superfluous (e.g., if he states, "I'll be indifferent if you make the orange sector 65% of the wheel"). At the same time, the wheel maintains the advantage of displaying both the event and its complement.

Plot each response as a point on a cumulative distribution (it is a good idea to number the points sequentially). An example is shown in Figure B-1. This will point out any inconsistencies and will also indicate gaps in the distribution that need one or more additional points. Do not, however, show this to the subject at this stage in the process because he may try to conform to a smooth curve.

Watch for shifts in responses that may indicate that the subject has remembered new information or shifted his thought process. The encoding process often causes the subject to change his original

156

INTERVIEWER __CHARLES STEEL__   SUBJECT __ED SMALLCASTLE__   DATE __6/14/78__

CUMULATIVE PROBABILITY (y-axis, 0.0 to 1.0)

SALES (x-axis, 0 to 3000+)

Numbers Indicate Time Sequence
X Indicates Responses to Odds Questions
O Indicates Responses to Wheel Questions

VARIABLE __THETA SALES AT MATURITY__
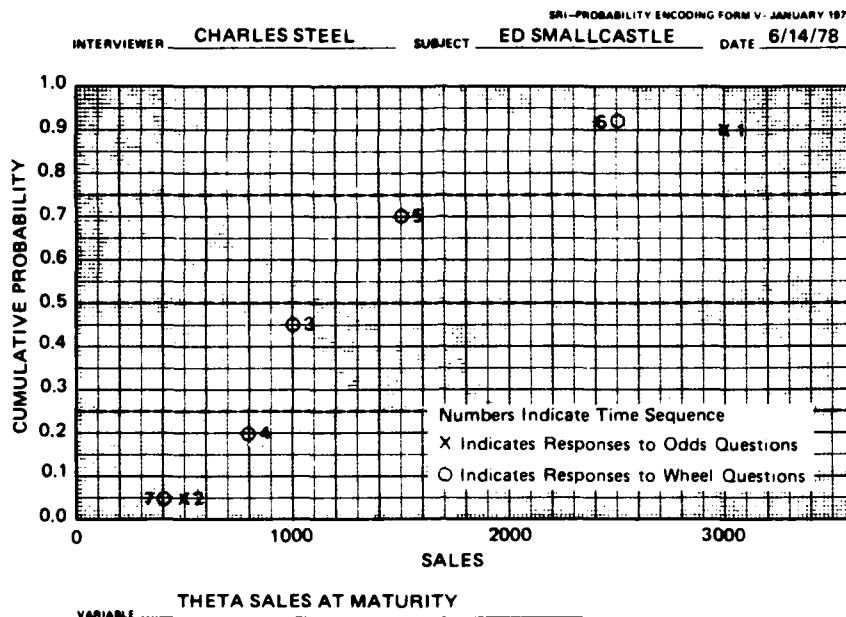
FIGURE B-1   GRAPHICAL REPRESENTATION OF RESPONSES

opinions as he is forced to think more clearly and fully about his knowledge. He will often verbalize this shift. The encoding objective is to capture his final judgment. Consistency with early responses should be discouraged.

Step 2: Elicitation of Median

Eliciting the median and the quartiles (see Step 3) is a way to ask check-questions. It is dangerous in that the responses leading to the quartiles are often anchored on the median. The results should be checked against those obtained with the wheel, and if there are any major differences, then those should be discussed with the subject.

The median is encoded by finding that value for which the subject finds it equally likely that the outcome will be below or above the value. Divide the range of possible outcomes into two parts and ask the subject which part he would prefer to bet on assuming the prizes were the same. The dividing point is changed to reduce the size (and

likelihood) of the part chosen, thereby increasing the size of the other part. The subject is then asked to choose between the two new parts. This procedure of changing the dividing point is continued until indifference is reached. The final dividing point then becomes the median--i.e., the 0.50 fractile. As in Step 1, the first two choices should be fairly simple: the first two dividing points should be well apart from where the median is expected to fall.

Example: Unit Manufacturing Cost for a New Machine

Question: Let us take the range of possible values of the manufacturing cost and divide it into two parts, cost being below $4,000 and cost being above $4,000 per unit. If you were offered the choice of betting on the cost coming out below $4,000 or above $4,000, which bet would you choose?

Answer: I would rather bet above $4,000.

Question: So, let us change the dividing point to $3,200. Now, would you bet on the cost being below $3,200 or the cost being above $3,200?

Answer: I would bet above $3,200.

Question: So we'll change the point to $3,600. Would you bet above or below?

Answer: Above.

Question: How about $3,800?

Answer: Then I'd bet below.

Question: How about $3,700?

Answer: That is about right. Roughly, the two intervals seem to be equally likely.

Remark: The final dividing point, $3,700, becomes the median --i.e., it corresponds to a probability of 50%.

Step 3: Elicitation of Quartiles

Divide the range of possible outcomes below the median into two parts and ask the subject which of the two parts he would like to bet on, assuming the prizes were the same. Adjust the dividing point until indifference is reached. Do the same for the range of outcomes above the median. The two points generated by this step are called the quartiles.

Example (continued):

Question: Consider the range of possible values of the cost that is less than $3,700. Let us divide that range into two parts, the cost being below $2,600 and the cost being between $2,600 and $3,700. If you were given the opportunity to bet on one of the two parts, which part would you choose?

Answer: I would bet on the cost being between $2,600 and $3,700.

Question: Let us change the dividing point to $3,300. Now, would you prefer to bet on the cost being below $3,300 or between $3,300 and $3,700?

Answer: I would bet below $3,300.

Question: What if we change the dividing point to $3,000?

Answer: Now I would bet above $3,000. About $3,100 would be the break-even point.

Question: Let's turn to outcomes above $3,700. If we divide the range above $3,700 into two parts, the cost being between $3,700 and $4,800, and the cost being above $4,800, which part would you prefer to bet on?

Answer: I would bet between $3,700 and $4,800.

Question: Let us change the dividing point to $4,500. Now which side would you bet on, between $3,700 and $4,500 or above $4,500?

Answer: Reduce the number a little and I will be indifferent. Make it $4,400.

It is easy to construct a checking question on the basis of the above responses because the interval between the two quartiles should have a probability of 0.50.

Example (continued):

Comment: You said earlier that betting below and above $3,700 was the same to you. In other terms, the two events of the cost being below $3,700 and the cost being above $3,700 were considered equally likely and therefore they could each be assigned a 50% probability. When we divided the interval below $3,700 into two equally likely parts, the dividing point was $3,100. That means that the probability of the cost being below $3,100 is half of 50%--that is, 25%. Similarly, the probability of the cost being above $4,400 is also 25% Do you follow? Ok, this means that the probability of the cost being between $3,100 and $4,400 is 25% + 25% = 50%. In other words, your responses imply that it is equally likely that the cost will be in the interval from $3,100 to $4,400 as it is that the cost will be outside that interval. Think about that; does it agree with your judgment?

Answer: Yes, I think so.

Remark: If the answer is negative, the questioning will have to continue as shown below.

Example (continued):

Answer: It seems more likely to be inside the interval than outside.

Question: Do you think you can adjust the interval to make the probability 50%?

Answer: I guess $3,200 to $4,300 would be about right.

Question: Would you say that $3,200 divides the range below $3,700 into two equally likely parts and that $4,300 does the same for the range above $3,700 units?

Answer: Yes.   [$3,200 and $4,300 are assigned the probability levels of 0.25 and 0.75.]

160

It may sometimes seem as if the 50% interval is very narrow and that there is reason to believe that the indifference points were reached by an insufficient adjustment from the starting point. In such cases, the interviewer might mention the adjustment principle (see Section III.2). Alternatively, he may prefer to point out the implications of the narrow interval.

Example (continued):

Comment: You have now defined an interval from $3,100 to
            $4,000, which in your judgment has a probability of
            50%. You said earlier when we discussed an intui-
            tive estimate that the cost is rather uncertain
            because of the new design features. Let me just
            mention to you that this interval covers only
            $1,100, which may seem a small number in relation
            to all possible outcomes. You may want to consider
            whether the interval should be more spread out.

Remark: The questions used to elicit the median and the quartiles have been phrased in terms of bets rather than in terms of which of two parts is more likely (which would be an equivalent formulation). The betting version is chosen to reduce the anchoring on the median, which is otherwise likely to occur. A subject may see the questions used to elicit the quartiles as trying to move him away from his original estimate.

Example (continued):

Question: If we divide the range below $3,700 (the median)
            into two parts, the cost being below $2,800 and the
            cost being between $2,800 and $3,700, which of the
            two parts do you think is more likely?

Answer: Above $2,800.

Question: How about the cost below $3,200 and the cost
            between $3,200 and $3,700?

Answer: Above $3,200.

Question: How about below $3,500 or between $3,500 and
            $3,700?

Answer: Between $3,500 and $3,700. I already told you that
            $3,700 was my best guess!

Step 4: New Information

The interviewer should be aware of the appearance of new information. In the middle of the encoding process, the subject may remember pieces of information that may bear on the outcome of the Quantity. Such information may invalidate all points so far in the process. When this occurs, it often is useful to probe the implications of the new information and then stop the session for awhile. This break is especially needed if a major amount of rework is necessary to further encode the Quantity.

Step 5: Change in Assumptions or Structure

The interviewer may find that the subject is introducing new assumptions or changing old ones. Such assumptions should be listed, and it should be checked whether earlier responses would change if these assumptions are included.

The subject's way of responding may indicate that he is reaching his answer by considering some conditions. If these conditions seem important to the problem, or if the subject is having difficulty in balancing the effects of the different conditions, then they should be made explicit and the encoding should be extended to be done for each of the different conditions.

Example: Cost of a New Subsystem

Question: You seem to find it difficult to answer the questions. Is it the questions or is it the problem itself that you find difficult to understand?

Answer: There's no problem with the question. I am just thinking what the effect will be if our vendor goes out of business.

Comment: Well, if your judgment about the cost of the subsystem depends on your vendor's failure, then maybe you would be more comfortable discussing the cost of the subsystem separately for the two conditions of your vendor going out of business and not going out of business.

Step 6: Fitting a Cumulative Distribution

After enough points have been encoded, a curve should be fitted to the points. An example is shown in Figure B-2.
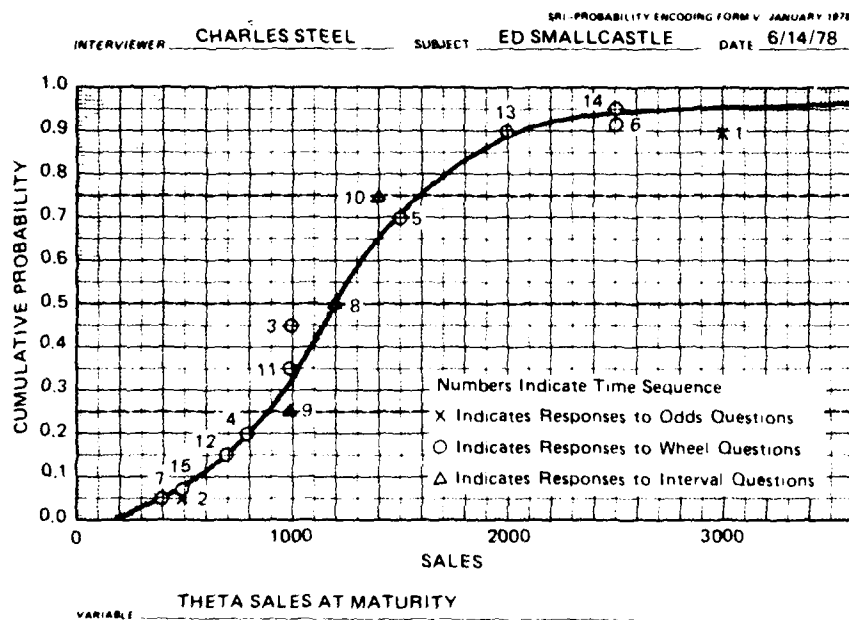
FIGURE B-2   EXAMPLE OF A CURVE FITTED TO RESPONSES

V VERIFYING

The judgments are tested to see if the subject really believes them. If the subject is not comfortable with the final distribution, it may be necessary to repeat some of the earlier steps of the interview process.

Step 1: Interpretation of Cumulative Distribution

Graphically representing the responses as points on a cumulative distribution and interpreting this distribution (perhaps in terms of a density function) provide an important test and feedback. The interviewer will naturally have to explain how the responses were plotted and how the fitted curve should be interpreted. An examination of the distribution itself cannot show whether or not the distribution agrees with the subject's judgment. However, it can show implications of the subject's responses and thereby provide feedback. *If some responses are not consistent with the subject's judgment, they will have to be modified.*

Example 1: Number of Tank Units in Use in a Specific Country Five Years from Now

Comment: Your responses to the various questions have been
            plotted as points in this diagram. [The
            interviewer shows Figure B-3(a) and explains the
            diagram and a few points.] This curve seems to fit
            the points reasonably well. However, let's take a
            look at the shape of the curve: first it rises,
            then levels off, and finally rises again. The
            probability of the number of tank units being
            between 600 and 1,000 is 0.53 - 0.36 = 0.17. At
            the same time, the interval from 400 to 600 has a
            probability of 0.27, and the interval from 1,000
            and 1,200 has a probability of 0.31. In other
            words, the number of tank units is judged more
            likely to be found around 500 or around 1,100 than
            around 800 which is a number between the other two.
            Do you see the implication of the curve, and do you
            have any comments?

165
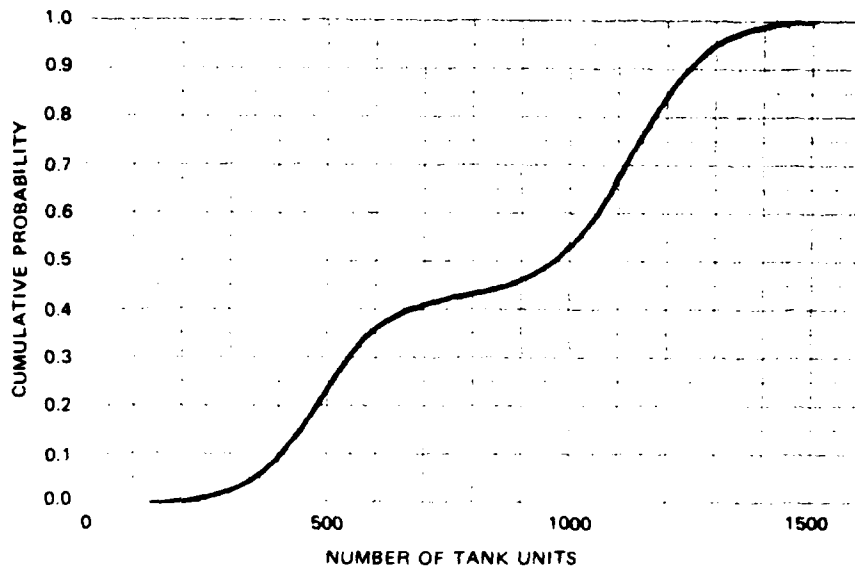
(a) CUMULATIVE DISTRIBUTION
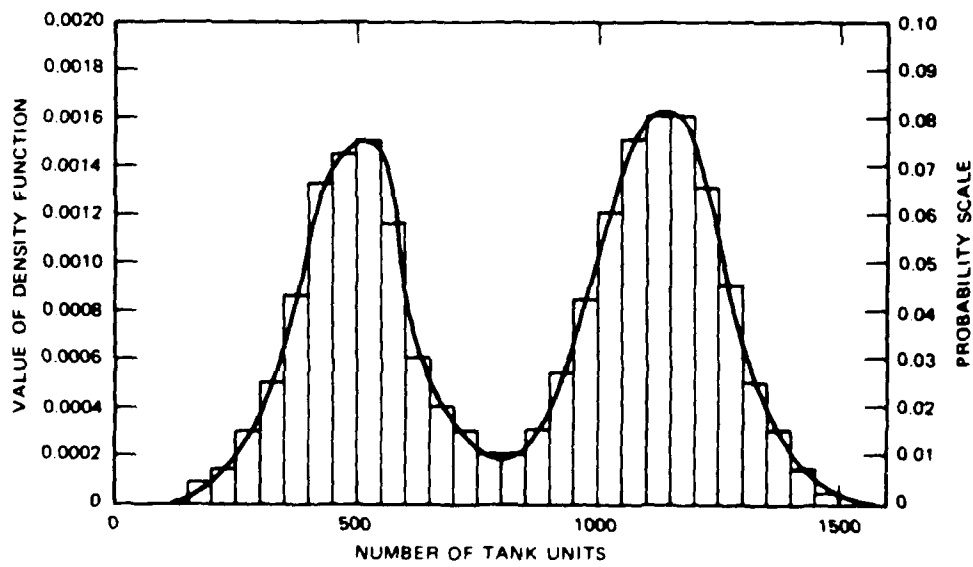
(b) APPROXIMATE DENSITY FUNCTION

FIGURE B-3   EXAMPLE OF DERIVATION OF A DENSITY FUNCTION FROM A
CUMULATIVE DISTRIBUTION

166

Answer 1: The way I see it, there are two cases: either the tank will work only in the desert, in which case the number will be around 500, or it will work in both the desert and the tropics, in which case the number will be around 1,100.

Answer 2: The curve looks funny. I don't see any reason it should be flat in between. I guess we'd better look at some of the points again.

Example 2: Average Time Between Failures for a New Component

Remark: [See Figure B-4.] The curve seems to be increasing rapidly, which means that the probability increases as the failure time increases, but then it drops abruptly after around 110 hours. For example, there is a 17% probability of the failure time falling between 105 and 110 hours at the same time as it is judged impossible that the failure time exceeds 110 hours.
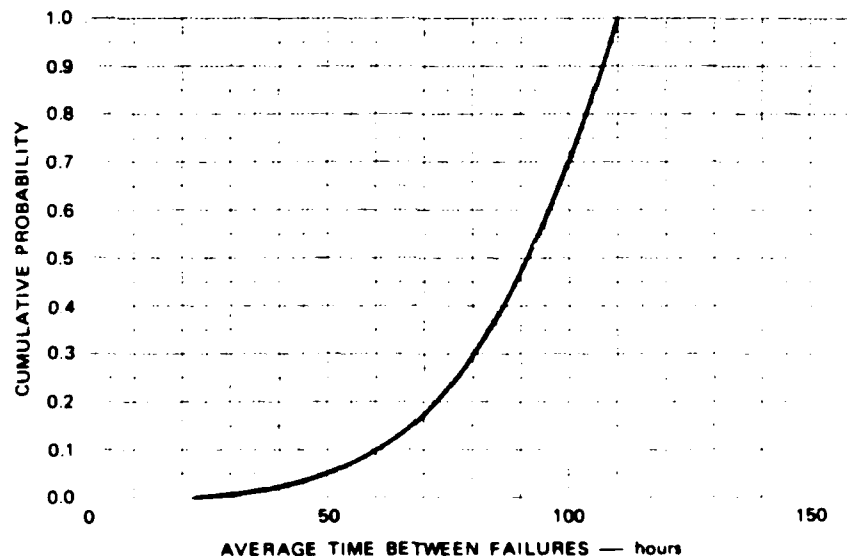
FIGURE B-4   EXAMPLE OF A CUMULATIVE DISTRIBUTION

Answer: You know the specifications state 110 hours between
           failures, so the research aims at driving the
           failure time below this level; at the same time,
           there is no incentive to improve the reliability
           beyond specifications. But I suppose it could
           happen that the failure rate actually ends up above
           110 hours. I'd give that a 5% probability.

Step 2: Drawing the Density Function

It may be easier to see some of the above arguments by sketching the density function. This is done in Figure B-3(b) for Example 1 above. The range of outcomes for the number of tank units was divided into intervals of 50 units in length. A bar is drawn above each interval, with the height of the bar equal to the probability of the interval as read off the cumulative distribution. The total area under all bars is 50 (the width of each interval). If all numbers on the vertical axis are reduced by a factor of 50, then the total area is one and the bar graph is an approximation of the density function. It is easy to fit a smooth curve through the bar graph to represent the density function. It is clear from Figure B-3(b) that the distribution is bimodal--i.e., it has two peaks.

Step 3: Construction of Bets from the Fitted Curve

The curve that was fitted in Section IV to responses is supposed to represent the subject's cumulative distribution. The interviewer should take a point on the curve, which thus shows the probability that the Quantity will not exceed a certain value, and construct two bets that should have equal value given that probability assignment. If the subject finds it difficult to choose--i.e., if he is indifferent between the two bets--this confirms the point on the curve. There should be a few such indifference responses before the process is ended. This provides the subject and the interviewer with confidence that the curve represents the subject's judgment.

Example: The Curve in Figure B.2

Question: [The interviewer can read from the curve that the probability of sales being below 1,200 units is 0.70.] Let's go back to the wheel and ask you a few more questions. Would you prefer to bet on orange at one spin of the wheel [the wheel is set at 70% orange] or on sales being below 1,200 units?

Answer: I'd say that they seem to be about the same.

Step 4: Conclusion of the Encoding Session

The analyst should show the subject the final curve and make sure that the subject views the curve as a fair representation of his judgment. When the subject indicates that the curve captures his belief concerning the Quantity and that he would be willing to base his own decisions on that curve, the session can be terminated. The session ends with the interviewer thanking the subject for his cooperation.

Appendix C

ENCODING INTERVIEW FORM

This appendix consists of a copy of an encoding interview form that is currently being used by decision analysts at SRI International. The form is used to assist in keeping a record of the interview. It is not intended to cover all possible developments of an encoding interview, but should include most quantitative questions.

The form should not be used unless the interviewer is familiar with the interview process. This process is described in Chapter 5 and in Appendix B.

ENCODING INTERVIEW FORM

Subject: _____

Interviewer: _____

Quantity: _____

Date: _____

MOTIVATING

The notes taken in the motivating phase
summarize the interviewer's observations
rather than answers to direct questions.

1. The subject's expertise with respect
   to the Quantity

2. Is the subject a stakeholder in the
   project?

3. Potential motivational biases

STRUCTURING

	Remarks
1. Definition of Quantity	
2. Approaches to the assessment of the Quantity	
3. List of relevant factors and information	

*When people say they have considered
everything they could think if, they
usually haven't (incompleteness
principle).*

4. List of assumptions

5. Redefinition of problem

6. Choice of a scale

CONDITIONING

1. Definition of a reference class

2. Assessment of the average and
   the variability for the reference
   class

   Average   _____

   Upper limit   _____

   Lower limit   _____

3. Less informed estimate (use if
   no reference class)

4. Intuitive estimate of the Quantity

5. Evaluation of predictability   Remarks

6. Discussion of extremeness of
   intuitive estimate (if applicable).

*When predictablility is limited, things
are rarely as good as one hopes, nor as
bad as one fears (moderation principle).*

7. Discussion of anchoring (if
   applicable)

*When you adjust a value, you rarely adjust
enough (adjustment principle).*

8. Discussion of scenarios for extreme outcomes

What would be a very high value?

How could such a value occur?

a.

b.

c.

d.

What is the probability of such a high value?

What would be a very low value?

How could such a value occur?

a.

b.

c.

d.

What is the probability of such a low value?

1. Elicitation of probabilities

Value	Probability of Outcome		
	less than or equal to value	or	greater than value
____	____		____
____	____		____
____	____		____
____	____		____
____	____		____
____	____		____
____	____		____
____	____		____
____	____		____
____	____		____
____	____		____
____	____		____
____	____		____

Answers are recorded on the enclosed graph.

187

2. Elicitation of median

| Value | Chosen bet (check one) | |
	Below	Above
____	____	____
____	____	____
____	____	____
____	____	____
____	____	____
____	____	____

Final Value _____

3. Elicitation of quartile

   a. Lower quartile

| Value | Chosen bet (check one) | |
	Below	Above
____	____	____
____	____	____
____	____	____
____	____	____
____	____	____
____	____	____

Final Value _____

189

b. Upper quartile

Value	Chosen bet (check one)	
	Below	Above
———	———	———
———	———	———
———	———	———
———	———	———
———	———	———
———	———	———

Final Value ———

c. Check Inside vs Outside quartiles

Adjusted values:

lower quartile ———

median ———

upper quartile ———

Median and quartiles are recorded on the graph.

Cumulative Probability

INTERVIEWER _____   SUBJECT _____   DATE _____

VARIABLE _____

COMMENT _____

1.0 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
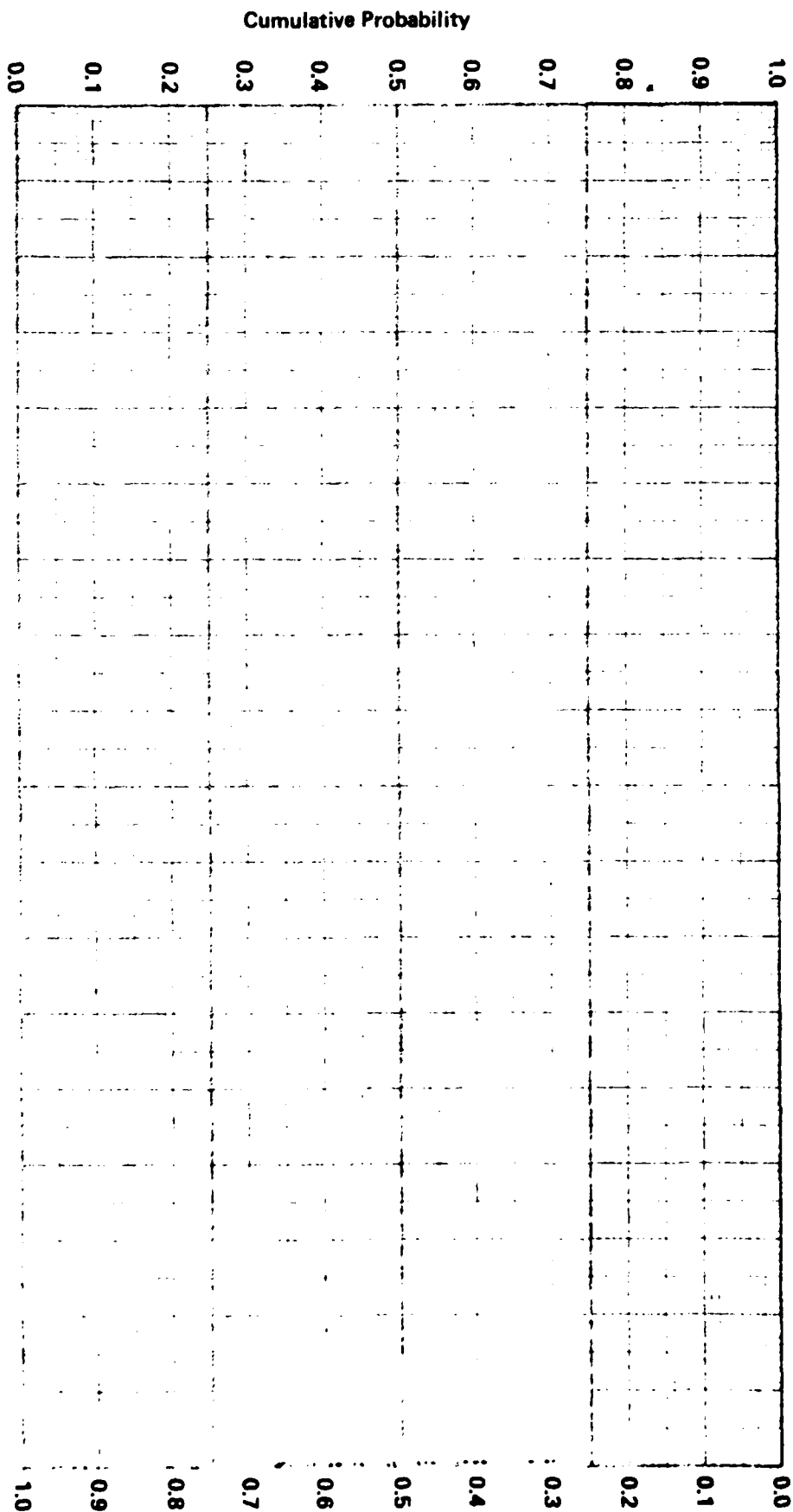
Appendix D

AN INTERACTIVE COMPUTER INTERVIEW

This appendix contains a sample session with the Probability Encoding Program (PEP) developed by the Decision Analysis Group at SRI International. This program is an interactive computer interview based on the interval technique for encoding. The subject is always given two or three intervals on which to bet and is asked to order the intervals in terms of his preferences. The program relies on an algorithm that finds the indifference point(s) by an iterative procedure. An opportunity always exists for the subject to bypass the iteration if he has decided on the indifference point. The program asks seven sets of questions, and the fractiles corresponding to the probability levels 1/6, 1/3, 1/2, 2/3, and 5/6 can be inferred from the answers. In fact, each fractile is encoded twice to provide a consistency check. The distribution is displayed in graphic and tabular form, and the subject is then made aware of any inconsistencies.

The easiest way to demonstrate the program is by means of a sample session. The following example required roughly 25 minutes and cost around $9.00. The instructions portion took up some of the time; an experienced user would find the length of the session to be between 10 and 15 minutes.

195

```
****************************************************************************
*  WELCOME TO THE DECISION ANALYSIS PROBABILITY ENCODING PROGRAM(PEP)  *
****************************************************************************
```

PLEASE TYPE YOUR NAME--<u>CARL-AXEL STAEL VON HOLSTEIN</u>
DO YOU NEED INSTRUCTIONS?   <u>YES</u>
THIS PROGRAM MEASURES YOUR PERSONAL JUDGEMENT ABOUT
AN UNCERTAIN QUANTITY. YOUR JUDGEMENT IS THEN INTERPRETED
AS A PROBABILITY DISTRIBUTION. KNOWLEDGE ABOUT PROBABILITY
WILL NOT NECESSARILY HELP YOU. THERE ARE ALSO NO RIGHT
OR WRONG ANSWERS BECAUSE YOUR JUDGEMENT IS UNIQUE.

SO RELAX AND TRY TO RESPOND WITH YOUR JUDGEMENT.

MOST OF THE QUESTIONS WILL BE IN A FORM WHERE YOU NEED
TO MAKE DECISIONS. YOU WILL BE GIVEN EITHER TWO OR
THREE NUMBERED RANGES. YOU WILL BE ASKED TO DECIDE
WHICH OF THE RANGES ARE THE MOST LIKELY.

ONE WAY TO ANSWER THE QUESTIONS IS TO ASK YOURSELF:
"WHICH OF THESE RANGES WOULD I PREFER TO BET ON?"

YOU ARE EXPECTED TO ORDER THE RANGES IN TERMS OF YOUR
PREFERENCES. FOR EXAMPLE, IF ONLY TWO RANGES ARE GIVEN,
YOU SIMPLY TYPE '12' OR '21'. IF YOU ARE GIVEN THREE
RANGES AND IF YOU PREFER 3 TO 2 AND 2 TO 1, YOU TYPE
'321'. IF YOU THINK 3 AND 2 ARE ABOUT EQUALLY LIKELY
BUT PREFER 3 AND 2 TO RANGE 1, TYPE '3=21'.

THE PROGRAM WILL ATTEMPT TO FIND RANGES THAT ARE EQUALLY
LIKELY IN YOUR JUDGEMENT. WHEN YOU HAVE REACHED THIS
POINT TYPE 'EQ' OR '1=2' OR '1=2=3'.

YOU ARE ALSO ALLOWED A NUMBER OF OTHER RESPONSES.

DO YOU WISH TO SEE THE ENTIRE LIST?   <u>YES</u>

RESPONSES   MEANING
---------   -------

 12,21   YOU ARE COMPARING TWO NUMBERED RANGES.
                     YOU PREFER THE FIRST NUMBERED RANGE
                     TO THE SECOND.

123,321,ETC...   YOU ARE COMPARING THREE NUMBERED
                     RANGES. YOU PREFER THE FIRST
                     NUMBERED RANGE TO THE SECOND AND THE
                     SECOND TO THE THIRD, ETC...

1=23,2=31,ETC...   YOU ARE AGAIN COMPARING THREE NUM-
                     BERED RANGES. THE EQUAL SIGN INDI-
                     CATES EQUALLY LIKELY RANGES.
                     THE RANGES ARE STILL ORDERED BY
                     DESIRABILITY OR PREFERENCE.

EQ,EQUAL   IN YOUR JUDGEMENT THE RANGES ARE EQUALLY
                     LIKELY. '1=2=3' AND '1=2' ARE EQUIVALENT
                     TO TYPING 'EQ'.

BAC,BACK UP   YOU WISH TO CHANGE A PREVIOUS RESPONSE.
                     YOU MAY BACK UP A MAXIMUM OF FIVE LINES
                     BY REFERENCING A LINE NUMBER. YOU MAY
                     ALSO RETURN TO THE START OF A SECTION OF
                     QUESTIONS BY TYPING THE SECTION LETTER.

TER,TERMINATE   YOU WANT TO JUMP OVER THE REST OF THE
                       QUESTIONS AND SEE THE OUTPUTS.

SKI,SKIP   .   YOU KNOW WHAT VALUES DIVIDE THE RANGES.
                       THE PROGRAM WILL REQUEST THAT YOU TYPE
                       THEM IN DIRECTLY. SIMPLY TYPE IN
                       YOUR VALUES WHEN THE '?' APPEARS.

DIF,DIFFERENCE   THE PROGRAM PRINTS THE DIFFERENCES
                       BETWEEN THE RANGE BOUNDRARIES.

SUP,SUPPRESS   THE PROGRAM STOPS PRINTING DIFFERENCES.

QUI,QUIT   QUIT ENTIRELY. YOU WANT TO JUMP OVER
                       THE REST OF THE QUESTIONS AND THE OUTPUT

WITH THIS INFORMATION, GO AHEAD AND TRY THE PROGRAM.
ALL OF YOUR RESPONSES MUST BE TERMINATED BY PRESSING THE
RETURN KEY.   FEEL FREE TO ASK FOR MORE INFORMATION AT ANY
TIME.   JUST TYPE THE WORD 'HELP'.

WHAT VARIABLE IS TO BE INVESTIGATED ?   <u>PROJECTED SALES</u>

DOES PROJECTED SALES HAVE UPPER AND LOWER LIMITS ?   <u>YES</u>
WHAT IS THE UPPER LIMIT ?   <u>500</u>
WHAT IS THE LOWER LIMIT ?   <u>100</u>

          [The program sets a limit to infinity if the
           limit is said not to exist.]

------------------**A**-------------------
            RANGE1   RANGE2
LINE#   FROM   TO/FROM   TO

  1   100.   300.   500.   <u>12</u>

  2   100.   195.   500.   <u>12</u>

  3   100.   147.   500.   <u>21</u>

  4   100.   170.   500.   <u>21</u>

  5   100.   183.   500.   <u>12</u>

  6   100.   176.   500.   <u>21</u>

  7   100.   180.   500.   <u>EQ</u>

      [180 is the 1 2 - fractile.]

```
--------------------**B**--------------------
          RANGE1    RANGE2    RANGE3
LINE#  FROM  TO/FROM  TO/FROM    TO

  1    100.    139.    322.     500.     HELP
```

THREE RANGES WERE JUST PRINTED OUT. YOU ARE EXPECTED
TO ORDER THESE THREE RANGES IN ACCORDANCE WITH THEIR
DESIRABILITY.
FOR EXAMPLE, IF YOU PREFER THE SECOND RANGE TO THE
FIRST RANGE AND THE FIRST RANGE TO THE THIRD, YOU
SHOULD RESPOND WITH '213'. YOU ARE ALLOWED TO USE
AN EQUAL SIGN TO INDICATE EQUALLY LIKELY RANGES,
SUCH AS '12=3'. THIS WOULD MEAN YOU CONSIDER THE
SECOND AND THIRD RANGES TO BE EQUALLY LIKELY,
BUT PREFER THE FIRST RANGE TO BOTH.

```
          RANGE1    RANGE2    RANGE3
LINE#  FROM  TO/FROM  TO/FROM    TO

  1    100.    139.    322.     500.     213

  2    100.    139.    248.     500.     213

  3    100.    139.    213.     500.     DIFFERENCES

  3    100.  (39.2)  139.  (73.8)  213.  (287.)  500.     231

  4    100.  (59.2)  159.  (53.8)  213.  (287.)  500.     1=23

  5    100.  (49.1)  149.  (47.0)  196.  (304.)  500.     21=3

  6    100.  (54.1)  154.  (33.7)  188.  (312.)  500.     321

  7    100.  (56.6)  157.  (35.3)  192.  (308.)  500.     E?
```

[157 and 192 are the 1-3- and 2-3-fractiles,
respectively.]

```
--------------------**C**--------------------
            RANGE1           RANGE2
LINE#   FROM       TO/FROM          TO

  1    100.  (28.3)  128.  (28.3)  157.     SUPPRESS

  1    100.    128.    157.     12

  2    100.    114.    157.     21

  3    100.    121.    157.     21

  4    100.    125.    157.     1=2
```

[125 is the 1-2-fractile since it divides an
interval with probability 1-2 into two equally
likely intervals.]

```
----------------**D**---------------
          RANGE1    RANGE2
LINE#  FROM  TO/FROM    TO

  1    192.    346.    500.    BACK
WHERE TO ?      C
```

[The subject realized that his indifference
point should be 127 for the range in Set C.]

```
----------------**C**---------------
          RANGE1    RANGE2
LINE#  FROM  TO/FROM    TO

  1    100.    128.    157.    SKIP

  1    100.   ?       127

  1    100.    127.    157.    EQ
```
```
----------------**D**---------------
          RANGE1    RANGE2
LINE#  FROM  TO/FROM    TO

  1    192.    346.    500.    12

  2    192.    265.    500.    SKI

  2    192.   ?       230

  2    192.    230.    500.    EQ
```

THIS ENDS THE FIRST QUESTION SET, A SHORTER SET FOLLOWS--PLEASE CONTINUE.

[230 is the 5 t fractile. The first four sets
of questions have established one value for
each of the five fractiles. The next three
sets are used for a consistency check. The
subject can terminate the interview at this
point (and anywhere else, too) and go directly
to the outputs.]

||| 1.0 |||2.8 |||2.5

||| |||3.2 ||| 2.2

||| 3.6

||| 4.0 ||| 2.0

||| 1.1 ||| 1.8

||| 1.25 ||| 1.4 ||| 1.6
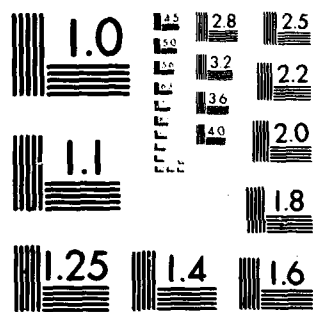
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

```
------------------**E**----------------------
       RANGE1     RANGE2     RANGE3
LINE#  FROM  TO/FROM  TO/FROM    TO

   1   100.    121.    161.    180.    213

   2   100.    121.    151.    180.    BACK
WHERE TO ?       1

   1   100.    121.    161.    180.    231

   2   100.    131.    161.    180.    231

   3   100.    137.    161.    180.    12=3

   4   100.    134.    151.    180.    31=2

   5   100.    135.    156.    180.    1=2=3
```

[The whole range was assigned probability 1/2
in Set A. 135 and 156 are therefore the 1/6-
and 1/3-fractiles.]

```
------------------**F**----------------------
       RANGE1     RANGE2
LINE#  FROM  TO/FROM    TO

   1   157.    174.    192.    21

   2   157.    183.    192.    12

   3   157.    179.    192.    EQ
```

[179 is a new value for the median (the 1/2-
fractile) since the limits of the range corre-
sponded to the probability levels 1/3 and 2/3
in Set B.]

DON'T GIVE UP--THIS IS THE LAST ONE
```
------------------**G**----------------------
       RANGE1     RANGE2     RANGE3
LINE#  FROM  TO/FROM  TO/FROM    TO

   1   180.    195.    341.    500.    SKI

   1   180.     ?     Q210
WHAT ?
210

   1   180.    210.     ?     290

   1   180.    210.    290.    500.    EQ
```

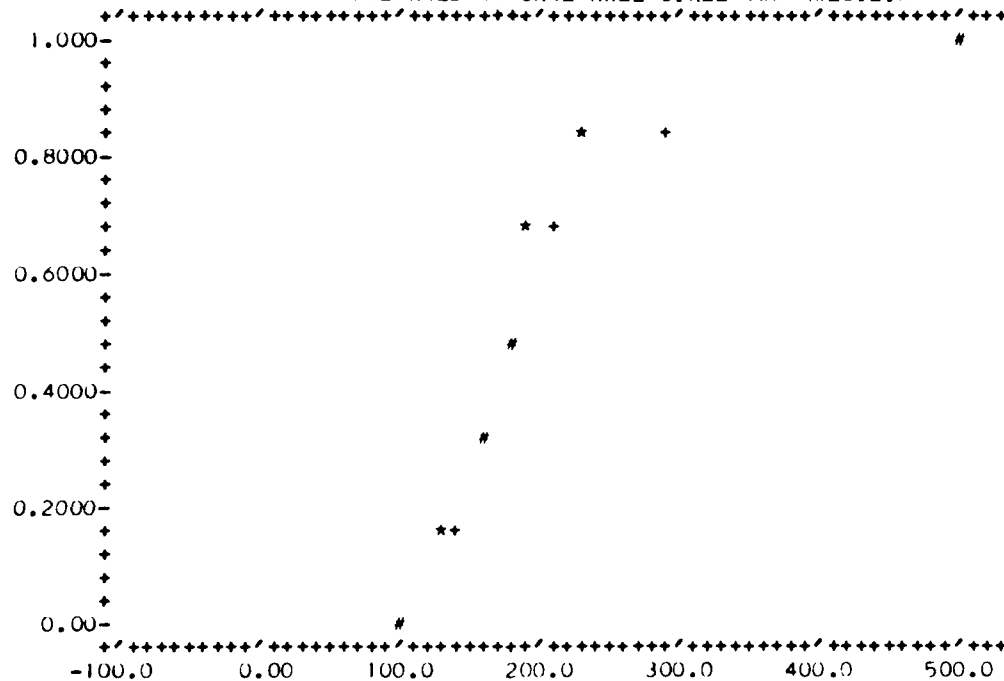[210 and 290 are new values for the 2/3- and
5/6-fractiles.]

```
*****************************************************************
*                         PEP : OUTPUT                          *
*****************************************************************
```

VARIABLE NAME:PROJECTED SALES   DATE 11/16/72

PREPARED BY CARL-AXEL STAEL VON HOLSTEIN

```
        +/++++++++/++++++++/++++++++/++++++++/++++++++/+++++++++/+++
 1.000- +                                                        #
        +
        +
        +
        +                                   *         +
0.8000- +
        +
        +
        +                              *  +
        +
0.6000- +
        +
        +                            #
        +
0.4000- +
        +                          #
        +
        +
        +
0.2000- +
        +                       *+
        +
        +
        +
 0.00-  +                     #
        +/++++++++/++++++++/++++++++/++++++++/++++++++/+++++++++/+++
        -100.0     0.00     100.0     200.0     300.0     400.0     500.0
```

CUMULATIVE PROBABILITY TABLE

PROBABILITY	THAT X IS LESS THAN	
	(****)	(++++)
1.00	500.	500.
.833	230.	290.
.667	192.	210.
.500	180.	179.
.333	157.	156.
.167	127.	135.
.000	100.	100.

[The subject can now see some inconsistencies,
especially at the high end. The 5/6-fractile was
first inferred to be 230 from Set D and later to
be 290 from Set F. The subject now has to re-
concile the inconsistencies, but that is done
outside of the program.]

201

CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

Director   2 copies
Advanced Research Projects Agency
Attention: Program Management Office
1400 Wilson Boulevard
Arlington, Virginia 22209

Office of Naval Research   3 copies
Attention: Code 455
800 North Quincy Street
Arlington, Virginia 22217

Defense Documentation Center   12 copies
Attention: DDC-TC
Cameron Station
Alexandria, Virginia 22314

DCASMA Baltimore Office   1 copy
Attention: Mr. K. Gerasim
300 East Joppa Road
Towson, Maryland 21204

Director   6 copies
Naval Research Laboratory
Attention: Code 2627
Washington, D.C. 20375

SUPPLEMENTAL DISTRIBUTION LIST
(Unclassified Technical Reports)

Department of Defense

Director of Net Assessment
Office of the Secretary of Defense
Attention: MAJ Robert G. Gough, USAF
The Pentagon, Room 3A930
Washington, DC 20301

Assistant Director (Net Technical Assessment)
Office of the Deputy Director of Defense
  Research and Engineering (Test and
  Evaluation)
The Pentagon, Room 3C125
Washington, DC 20301

Director, Defense Advanced Research
  Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, ARPA Regional Office (Europe)
Headquarters, U.S. European Command
APO New York 09128

Director, ARPA Regional Office (Pacific)
Staff CINCPAC, Box 13
APO San Francisco 96610

Dr. Don Hirta
Defense Systems Management School
Building 202
Ft. Belvoir, VA 22060

Chairman, Department of Curriculum
  Development
National War College
Ft. McNair, 4th and P Streets, SW
Washington, DC 20319

Defense Intelligence School
Attention: Professor Douglas E. Hunter
Washington, DC 20374

Vice Director for Production
Management Office (Special Actions)
Defense Intelligence Agency
Room 1E863, The Pentagon
Washington, DC 20301

Command and Control Technical Center
Defense Communications Agency
Attention: Mr. John D. Hwang
Washington, DC 20301

Department of the Navy

Office of the Chief of Naval Operations
  (OP-951)
Washington, DC 20450

Office of Naval Research
Assistant Chief for Technology (Code 200)
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

205

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Director, ONR Branch Office
Attention: Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
Attention: Dr. M. Bertin
American Embassy - Room A-407
APO San Francisco 96503

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, DC 20380

Headquarters, Naval Material Command
(Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Head, Human Factors Division
Naval Electronics Laboratory Center
Attention: Mr. Richard Coburn
San Diego, CA 92152

Dean of Research Administration
Naval Postgraduate School
Attention: Patrick C. Parker
Monterey, CA 93940

Superintendent
Naval Postgraduate School
Attention: R. J. Roland, (Code 52R1)
C^3 Curriculum
Monterey, CA 93940

Naval Personnel Research and Development
Center (Code 305)
Attention: LCDR O'Bar
San Diego, CA 92152

Navy Personnel Research and Development
Center
Manned Systems Design (Code 311)
Attention: Dr. Fred Muckler
San Diego, CA 92152

Naval Training Equipment Center
Human Factors Department (Code N215)
Orlando, FL 32813

Naval Training Equipment Center
Training Analysis and Evaluation Group
(Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Director, Center for Advanced Research
Naval War College
Attention: Professor C. Lewis
Newport, RI 02840

Naval Research Laboratory
Communications Sciences Division (Code 5403)
Attention: Dr. John Shore
Washington, DC 20375

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Chief, Intelligence Division
Marine Corps Development Center
Quantico, VA 22134

Department of the Army

Alan H. Curry
Operations and Management Science Division
U.S. Army Institute for Research in Manage-
  ment Information and Computer Science
730 Peachtree St., N.E. (Suite 900)
Atlanta, Georgia 30308

Deputy Under Secretary of the Army
(Operations Research)
The Pentagon, Room 2E621
Washington, DC 20310

Director, Army Library
Army Studies (ASDIRS)
The Pentagon, Room 1A534
Washington, DC 20310

U.S. Army Research Institute
Organizations and Systems Research Laboratory
Attention: Dr. Edgar M. Johnson
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and Systems
  Research Laboratory
U.S. Army Institute for the Behavioral
  and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director, U.S. Army Concepts
  Analysis Agency
8120 Woodmont Avenue
Bethesda, MD 20014

Director, Strategic Studies Institute
U.S. Army Combat Developments Command
Carlisle Barracks, PA 17013

Commandant, Army Logistics Management Center
Attention: DRXMC-LS-SCAD (ORSA)
Ft. Lee, VA 23801

Department of Engineering
United States Military Academy
Attention: COL A. F. Grum
West Point, NY 10996

Marine Corps Representative
U.S. Army War College
Carlisle Barracks, PA 17013

Chief, Studies and Analysis Office
Headquarters, Army Training and Doctrine
  Command
Ft. Monroe, VA 23351

Commander, U.S. Army Research Office
  (Durham)
Box CM, Duke Station
Durham, NC 27706

Department of the Air Force

Assistant for Requirements Development
  and Acquisition Programs
Office of the Deputy Chief of Staff for
  Research and Development
The Pentagon, Room 4C331
Washington, DC 20330

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Commandant, Air University
Maxwell AFB, AL 36112

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Deputy Chief of Staff, Plans, and
  Operations
Directorate of Concepts (AR/XOCCC)
Attention: Major R. Linhard
The Pentagon, Room 4D 1047
Washington, DC 20330

Director, Advanced Systems Division
  (AFHRL/AS)
Attention: Dr. Gordon Eckstrand
Wright-Patterson AFB, OH 45433

Commander, Rome Air Development Center
Attention: Mr. John Atkinson
Griffis AFB
Rome, NY 13440

207

IRD, Rome Air Development Center
Attention: Mr. Frederic A. Dion
Griffis AFB
Rome, NY 13440

HQS Tactical Air Command
Attention: LTCOL David Dianich
Langley AFB, VA 23665

Other Government Agencies

Chief, Strategic Evaluation Center
Central Intelligence Agency
Headquarters, Room 2G24
Washington, DC 20505

Director, Center for the Study of
  Intelligence
Central Intelligence Agency
Attention: Mr. Dean Moor
Washington, DC 20505

Mr. Richard Heuer
Methods & Forecasting Division
Office of Regional and Political Analysis
Central Intelligence Agency
Washington, DC 20505

Office of Life Sciences
Headquarters, National Aeronautics and
  Space Administration
Attention: Dr. Stanley Deutsch
600 Independence Avenue
Washington, DC 20546

Other Institutions

Department of Psychology
The Johns Hopkins University
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Stanford University
Attention: Dr. R. A. Howard
Stanford, CA 94305

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A. D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Brunel University
Attention: Dr. Lawrence D. Phillips
Uxbridge, Middlesex UB8 3PH
England

Decision Analysis Group
Stanford Research Institute
Attention: Dr. Miley W. Merkhofer
Menlo Park, CA 94025

Decision Research
1201 Oak Street
Eugene, OR 97401

Department of Psychology
University of Washington
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Department of Electrical and Computer
  Engineering
University of Michigan
Attention: Professor Kan Chen
Ann Arbor, MI 94135

Department of Government and Politics
University of Maryland
Attention: Dr. Davis B. Bobrow
College Park, MD 20747

Department of Psychology
Hebrew University
Attention: Dr. Amos Tversky
Jerusalem, Israel

208

Dr. Andrew P. Sage
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901

Professor Raymond Tanter
Political Science Department
The University of Michigan
Ann Arbor, MI 48109

Professor Howard Raiffa
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

Department of Psychology
University of Oklahoma
Attention: Dr. Charles Gettys
455 West Lindsey
Dale Hall Tower
Norman, OK 73069

Institute of Behavioral Science #3
University of Colorado
Attention: Dr. Kenneth Hammond
Room 201
Boulder, Colorado 80309